ANALYSIS OF CONTEXT-DEPENDENT MODULATION OF EPIGENETIC GENE ACTIVITIES IN *CAENORHABDITIS ELEGANS*

by

Kyle "Ky" C. Sha

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
August 2006

ABSTRACT

We initiated a screen for *trans*-acting factors that modulate expression of a transgene reporter in *C. elegans*. In the process, we fortuitously generated transgene derivatives that exhibited an imprinting effect. Imprinting is the non-equivalence of reciprocal matings. From a single progenitor line carrying an extra-chromosomal *unc-54::gfp* transgene array, we generated three independent autosomal integrations of the *unc-54::gfp* transgene. The progenitor line, two of its three integrated derivatives, and a non-related *unc-119::gfp* fusion exhibit an imprinting effect: single-generation transmission of these transgenes through the male germline results in approximately 1.5-2.0 fold greater expression than transmission through the female germline. There is a detectable resetting of the imprint after passage through the opposite germline for a single generation, indicating that the imprinted status of the transgenes is reversible. In cases where the transgene is maintained in either the oocyte lineage or sperm lineage for multiple, consecutive generations, a full reset requires passage through the opposite germline for several generations. Taken together, our results indicate that *C. elegans* has the ability to imprint chromosomes and that differences in the cell and/or molecular biology of oogenesis and spermatogenesis are manifest in an imprint that can persist in both somatic and germline gene expression for multiple generations.

To gain insight into *C. elegans* chromatin, we expressed the *E. coli* dam methyltransferase in *C. elegans* muscle cells in order to map susceptible regions of the genome to DNA modification. Dam methyltransferase catalyzes the transfer of a methyl from S-adenosyl-methionine to adenine in the target sequence GATC. Using restriction enzymes that are sensitive to the methylation status of GATC, we mapped 309 dam

methyltransferase sites out of a potential total of 269,000 sites distributed more or less uniformly throughout the *C. elegans* genome. Our preliminary analysis indicated that targets of DAM are distributed uniformly throughout the entire genome, without any apparent bias towards specific chromosomes or genomic regions. SAGE analysis revealed that dam methyltransferase does not have any bias towards muscle-specific genes. Thus, our experimental system is a potentially useful tool for investigating genome-wide chromatin accessibility from within muscle tissue.

Thesis advisor: Prof. Andrew Fire

Thesis Reader: Prof. Jun Kelly Liu (Cornell University)

Thesis Defense Committee: Prof. Victor Corces (JHU), Prof. Mark van Doren (JHU), Prof. Karen Beemon (JHU)

— To my mom —

Who devoted her life to her children

— To my twin baby sisters Jackie and Krisstina —

Who taught me so much about life and, in the process, made me a better person

— To my three uncles —

Who were father-figures to me during my childhood

— To Dr. M.S. Smith —

Who taught me so much about myself and about life

— To Mrs. Hersey —

Who was a role model to me during my childhood

PREFACE

This dissertation has been six years in the making. Along the way, there have been individuals who have contributed one way or another, big or small, to my graduate education. First and foremost, I would like to express my gratitude to my mentor, Dr. Andrew Fire. I have learned that to be a student of Andy Fire is to be more than a student of science, but also to be a student of the virtues that define a good human being. Rarely does one encounter a person who is so intelligent, yet extremely humble and kind at heart. I am very lucky to have had him as my mentor.

I would like to acknowledge present and past members of the Fire lab who have offered helpful discussions and/or reagents: Daniel Blanchard, Chaya Krishna, Poornima Parameswaran, Jonathan Gent, and Steve Johnson. I thank the following Fire lab alumni: Jenny Hsieh (UT, Austin), whose work in the Fire lab I referred to frequently, Judy Yanowitz (CIW) for helpful discussions on the projects, Susan Parrish (NIH) for being my mentor during my rotation through the Fire lab, Steve Kostas, Bill Kelly (Emory University), and a special thank you to Kelly Liu (Cornell University) for critical reading of my dissertation.

I would also like to thank individuals not in the Fire lab who contributed to my graduate career at Johns Hopkins: Prof. Mariano Labrador (Univ. Tennessee, Memphis), who was my mentor during my rotation through the Corces lab, Drs. Kristen Nastase and Mimi Cheng, who were my mentors during my rotation through the Shearn lab, and Dr. Stephen King, who was my mentor during my rotation through the Schroer lab. The following faculty members have monitored my progress and made sure I was on track with my research:

Thesis Defense Committee	Thesis Committee
Prof. Victor Corces (Johns Hopkins)	Prof. Victor Corces
Prof. Mark van Doren (Johns Hopkins)	Prof. Mark van Doren
Prof. Karen Beemon (Johns Hopkins)	Prof. Alex Bortvin (CIW)

I thank the following faculty members who served on my qualifying (oral) examination committee: Prof. Judith Bender (JHU), Prof. Douglas Koshland (CIW), Prof. Allen Shearn (JHU), Prof. Trina Schroer (JHU), and Prof. Yixian Zhang (CIW). I would like to thank Joan Miller, administrative assistant to the Biology Department. From my very first day at Johns Hopkins to my very last (and even after I moved to Stanford University with the Fire lab), Joan was always available to assist in all administrative matters.

I would like to thank faculty members not affiliated with Johns Hopkins University; yet who have contributed one way or another to my dissertation: Prof. Jonathan Hodgkins (Oxford University, UK), for helpful discussions on the imprinting project; Prof. Anne Villeneuve (Stanford University), for critical reading of the imprinting manuscript and her numerous suggestions and helpful discussions on the

imprinting project; Prof. Emma Whitelaw (University of Sydney, Australia) for reading of the imprinting manuscript.

Finally, I thank the following institutions and centers for academic, financial, facilities, and/or reagents support:

- National Institutes of Health (NIH) - Bethesda, MD
- Johns Hopkins University, Biology Department - Baltimore, MD
- Carnegie Institution of Washington, Dept. of Embryology - Baltimore, MD
- Stanford University, Depts. of Pathology and Genetics - Stanford, CA
- *C. elegans* Genetics Center - Minneapolis, MN

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

In 1744 Linnaeus, the founder of the modern biological system of classification, described a deviant specimen of the common toad-flax (*Linaria vulgaris*) that he acquired from another Swedish botanist. The specimen was radially symmetric rather than the bilaterally symmetric shape of the wildtype flower (GUSTAFSSON 1979). Two hundred fifty-five years later, a group of geneticists solved the mystery of Linnaeus' odd discovery: The *Lcyc* gene in the mutant, responsible for floral symmetry in *Linaria vulgaris*, had been inactivated by DNA hypermethylation (CUBAS *et al*. 1999). There was no change in DNA sequence between the wildtype and mutant alleles. This finding was rather unexpected; first because it was an epimutation, rather than a genetic change, that caused the phenotype. Secondly, the observation that the epimutation could be meiotically stable for so long was also surprising.

Non-Mendelian genetic phenomena such as imprinting, paramutation, and X inactivation, have been known to geneticists for decades. Such phenomena were described, but the mechanisms behind them remained elusive, partly due to lack of technology, and partly due to the challenges inherent in the study of epigenetic phenomena. These include non-Mendelian segregation, variable penetrance between independent lines, and the tendency for revertants to appear in the population. It was only in the past two or so decades that the molecular mechanisms behind epigenetic variation began to be unraveled. Today, the study of epigenetic phenomena is widespread, and the mechanisms uncovered provide new paradigms for understanding how cells reversibly regulate gene action without causing a change in the primary sequence of the gene.

Insight into epigenetic mechanisms has allowed researchers to explain phenomena that would otherwise be extremely difficult to characterize using Mendelian genetics,

such as why one twin can be consistently more susceptible to a certain pathology than his/her identical twin (raised in the same environment), why animal cloning is so easy in principle, but extremely challenging technically (and why the clone is not simply a replica of the DNA donor), why it is that the harder one tries to express a transgene (i.e. by providing additional copies), the less expression one gets, and why in paramutation a phenotypic variation can persist when the responsible locus is segregated away. Increasingly, epigenetics is being implicated in developmental disorders and imprinting-related diseases (JIANG *et al*. 2004; MARTIN *et al*. 2005), cancer (BAYLIN and OHM 2006), and even mental illnesses (PETRONIS *et al*. 1999). Ultimately, to study epigenetics *is* to study genetics. The former does not supercede the latter, but complements it. An understanding of epigenetics adds another layer of understanding to genetic phenomena.

To understand the fundamental mechanisms by which epigenetics operates, it is necessary to understand chromatin. We do not know if all epigenetic phenomena operate through chromatin, but numerous studies of epigenetic processes have thus far frequently pointed to a chromatin explanation. Changes in chromatin states, mediated by DNA methylation, histone modifications, metastable differences in the distribution of associated protein factors (i.e. Polycomb and trithorax proteins), and/or non-coding RNAs, are the fundamental events by which cells keep the "on" or "off" state of a gene through mitosis or meiosis. In this chapter, we discuss genomic imprinting, with an emphasis on its mechanisms. We focus primarily on mammals, the system in which imprinting is best understood. Where relevant, we also discuss examples from other systems, either to augment or to contrast with our current understanding of imprinting mechanisms in mammals. The manner by which chromatin modulates the imprinting

process (its establishment, its maintenance, its erasure) can be abstracted to numerous other epigenetic phenomena.

Imprinting: background

Imprinting is the non-equivalence of reciprocal matings. In diploid organisms, every gene has two copies, one contributed by the female parent (the maternal allele) and one by the male parent (the paternal allele). In most model systems, the vast majority of genes can be expressed from either the maternal or the paternal allele. However, for a small subset of genes (approximately 80 thus far identified in humans and slightly more in mice; http://www.mgu.har.mrc.ac.uk/research/imprinting/imprin-intro.html), there are known differences observed when the maternal and paternal alleles are measured separately. Failure to express the proper allele in the correct temporal and/or spatial requirement can result in mental and/or physical disorders (if the effect is mild) or abortion of the fetus in severe cases. Imprinting, then, violates the Mendelian principle that the two segregating alleles of a gene are equivalent. Geneticists recognize imprinted genes by the non-equivalence of reciprocal matings. But how does an organism recognize which allele of an imprinted gene came from which parent? The primary sequence only plays a partial role. Studies of imprinting in multiple organisms all implicate the involvement of chromatin. The chromatin state of imprinted paternal and maternal alleles are not identical. Somehow, the progeny is able to decipher differences in chromatin state between the paternal and maternal alleles.

Imprinting mechanisms

genomic context: What distinguishes an imprinted gene from a non-imprinted gene? There is no one single characteristic that distinguishes the two. Rather, it is a list of

parameters that vary among imprinted genes. Although not universal, many imprinted genes share certain common genomic features. One study indicated that imprinted genes contain fewer and smaller introns (HURST *et al.* 1996). However, many exceptions to this observation were found. Neumann *et al.* observed that imprinted genes tend to exhibit some degree of repetitive sequences (NEUMANN *et al.* 1995). This observation has been corroborated by comparative genomics studies that indicate a high content of retrotransposable elements in imprinted regions, and of the approximately 80 genes known to be imprinted in mammals, over 23 have tandemly repeated sequences in them (WALTER *et al.* 2006). Mary Lyon, noting the high abundance of repetitive elements on the mammalian X chromosome, suggested that perhaps repetitive elements facilitate the initiation and spreading of a heterochromatic state during dosage compensation (LYON 1998). Other investigators have since adapted Lyon's hypothesis to imprinted loci. Although it is generally the case that imprinted genes contain repetitive elements, the actual role such repetitive elements play in the imprinting process is still not clear. For example, the *Impact* gene is imprinted in mouse, rat, and rabbit (OKAMURA *et al.* 2005). Tandem repeats within the mouse and rat *Impact* gene are methylated in a parent-of-origin specific manner. The rabbit gene lacks the tandem repeats, but is apparently still imprinted (OKAMURA *et al.* 2005). Other known examples of imprinted genes that do not involve nearby repetitive elements include *MEDEA* in *Arabidopsis* (SPILLANE *et al.* 2004) and the non-coding RNA *Kcnq1ot1* in mouse (MANCINI-DINARDO *et al.* 2006).

In mammals, approximately 80% of imprinted genes exist in clusters or in close proximity to each other (REIK and WALTER 2001; VERONA *et al.* 2003). It is believed that differential cytosine methylation of the imprint control regions (ICRs) contributes to

coordinated expression of imprinted genes (AINSCOUGH *et al*. 1997; REINHART *et al*. 2002; THORVALDSEN *et al*. 1998; WUTZ *et al*. 1997). For example, the KCNQ1 gene in humans contains an imprint control region called ICR2 (MANCINI-DINARDO *et al*. 2003). In addition to controlling the imprinting of KCNQ1, ICR2 also controls the imprinting of five other genes in the cluster, including CDKN1C, ACL2, PHLDA2, TSSC4, and SLC22A1L (FITZPATRICK *et al*. 2002; HORIKE *et al*. 2000).

Further evidence linking genomic context and imprinting of genes comes from studies in *Drosophila*. Imprinting of endogenous genes has not been reported for *Drosophila*. Nevertheless, certain chromosomal translocations have conferred a parent-of-origin effect to previously non-imprinted endogenous genes (LLOYD 2000). It has even been shown that all genes on the translocated cluster can acquire a parent-of-origin-dependent expression pattern. Translocations that confer imprinting properties to genes that were previously non-imprinted often occur in or next to heterochromatic regions of the genome.

***cis*-acting elements:** Extensive studies also suggest contribution from regulatory signals within or in proximity to imprinted genes. Called imprint control regions (ICR) or differentially methylated regions/domains (DMR/D), such *cis*-acting elements are often the sites of DNA methylation establishment and/or maintenance in a manner dependent upon the parent-of-origin (HOLMES and SOLOWAY 2006). DMRs are relatively rich in CpG dinucleotides, but the CpG content is generally less than that of CpG islands (KOBAYASHI *et al*. 2006). Also, repetitive elements found in imprinted clusters are usually located within DMRs (WALTER *et al*. 2006). Depending on the methylation

status, the *cis*-acting sequences recruit different regulatory factors, such as MBD proteins, which contain a domain that recognizes and binds methylated DNA.

Certain DMRs can also function as insulators. An insulator has two properties (LABRADOR and CORCES 2002): (1) when placed between an enhancer and its target promoter, the insulator prevents activation of the promoter, (2) when placed upstream and/or downstream of a transgene, the insulators protect the transgene from position effects. One of the better-studied paradigms of a *cis*-acting element with both DMR and insulator roles is the DMR at the *H19*/*Igf2* imprinted region (Figure 1.1). *Igf2* and *H19* are reciprocally imprinted genes, with *Igf2* being expressed from the paternal allele and *H19* from the maternal allele (RACHMILEWITZ *et al*. 1992). The DMR of the maternal allele is not methylated. As a result, CTCF is able to bind the DMR and prevent *Igf2* promoter activation by a downstream enhancer. In the paternal allele, the methylated DMR prevents CTCF binding, and as a result the downstream enhancer is able to activate transcription of *Igf2* (BELL and FELSENFELD 2000; HARK *et al*. 2000; KANDURI *et al*. 2000; SZABO *et al*. 2000). Insulators can be swapped between different organisms and still retain buffering function, suggesting an evolutionarily conserved mechanism by which they operate.

Whatever the mechanism(s) of imprinting control by DMRs, transgene experiments provide strong evidence that they can confer imprinting status to non-imprinted sequences when placed under their regulation. For example, *RSVIgmyc* is a mouse transgene created from the Rous sarcoma virus LTR plus fragments from the Ig heavy chain and *c-myc* gene (SWAIN *et al*. 1987). During the course of characterization of *RSVIgmyc*, Swain *et al*. fortuitously discovered that *RSVIgmyc* was expressed in a

parent-of-origin manner. Thus, an imprinted transgene had been created from non-imprinted sources. Reinhart *et al*. created a non-imprinted derivative of *RSVIgmyc* by mapping and deleting its DMR (REINHART *et al*. 2002). Imprinting of the non-imprinted derivative was restored by substitution of a DMR from the *Igf2*/*H19* locus (REINHART *et al*. 2002). In another similar experiment, the *H19* DMR was able to confer imprinting status to the normally non-imprinted β-globin locus (TANIMOTO *et al*. 2005). This and other DMR swapping experiments suggest the presence of shared features between DMRs of different imprinted loci.

covalent modifications: What additional marks, besides *cis*-acting regulatory elements and genomic context, might an organism use to distinguish maternal from paternal alleles? So far as is known in all organisms examined, imprinting is invariably associated with 5-methyl-cytosine methylation (in those organisms that have m5C methylation). This is especially relevant in organisms that use CpG islands as promoter elements, such as mammals and plants. Differences in regional and/or degree of CpG island methylation may provide additional marks to distinguish between the maternal and paternal allele. For example, the ICR2 of KCNQ1, mentioned above, is a CpG island. The maternal ICR2 is completely methylated whereas the paternal allele is devoid of any detectable methylation (BEATTY *et al*. 2006).

DNA methylation is also involved in imprinting in insects. Due to the divergent biology between insects and mammals, there are differences in the relationship between DNA methylation and imprinting in these two groups. Insect imprinting is best characterized in the sciarid flies and mealybugs (coccid insects). Unlike mammals, where imprinting occurs at certain sparse clusters of genes, imprinting in mealybugs occurs by

heterochromatization and elimination of the entire paternal genome (BONGIORNI *et al.* 1999; KHOSLA *et al.* 2006). In sciara flies, there is a report of 5-methyl-cytosine methylation (EASTMAN *et al.* 1980), but a relationship between DNA methylation and imprinting has not been determined for Sciara.

In mammals, the continual methylation status of an allele requires two sets of DNA methyltransferases. During gametogenesis in mammals, *de novo* methyltransferases *Dnmt3a* and *Dnmt3L* methylate DMRs in a gamete-specific manner (BOURC'HIS *et al.* 2001; KANEDA *et al.* 2004). The action of *de novo* methylation, then, result in distinct DNA methylation marks between the two parental alleles. After fertilization, maintenance methyltransferases such as *Dnmt1* maintain the methylation state (GOLL and BESTOR 2005; HOWELL *et al.* 2001). It is generally the case that methylation of an allele is associated with lower activity of that allele. In plants, maintenance methylation, but not *de novo* methylation, is required for imprinting (JULLIEN *et al.* 2006; SCOTT and SPIELMAN 2004; VIELLE-CALZADA *et al.* 1999). This is apparently due to the different mechanism by which mammals and plants imprint genes (discussed below in "selective inactivation versus selective activation" section).

How might ICRs coordinate control of multiple genes within a given cluster? Turker and Bestor originally proposed a model in which DNA methylation is initially established at an ICR and subsequently spread to surrounding regions (TURKER and BESTOR 1997). A more recent model posits that ICRs can modulate long-range chromatin interactions through looping (LOPES *et al.* 2003; MURRELL *et al.* 2004). Both models appear to be supported by experimental evidence.

Despite the ubiquity of DNA methylation in imprinting, many investigators concede, however, that it is not known whether cytosine methylation is the cause or merely the manifestation of the imprinting process. In some cases, DNA methylation appears to be essential for proper expression of imprinted genes. For example, aberrant DNA methylation has been linked to a number of imprinting-related disorders in humans, apparently due to improper spatial and/or temporal allele-specific expression (PAULSEN and FERGUSON-SMITH 2001; ROBERTSON 2005). In *Arabidopsis*, maintenance of DNA methylation is required for the proper imprinted expression of *FWA* and *FIS2* (JULLIEN *et al*. 2006; KINOSHITA *et al*. 2004; SCOTT and SPIELMAN 2004). However, DNA methylation appears to play only a partial role in regulating imprinting of the *Arabidopsis MEDEA* gene (GEHRING *et al*. 2006; JULLIEN *et al*. 2006). Examples of a non-essential role for DNA methylation in genomic imprinting comes from *Drosophila* and *C. elegans*. Although cytosine methylation has been detected in *Drosophila* (KUNERT *et al*. 2003; LYKO *et al*. 2000), its role in the ability of *Drosophila* to imprint endogenous genes in certain cases has not been reported. DNA methylation has not been reported for *C. elegans* (GUTIERREZ and SOMMER 2004; HODGKIN 1994; SIMPSON *et al*. 1986), yet this species can apparently imprint a subset of transgenes (SHA and FIRE 2005) and establish certain gamete-specific histone marks on the X chromosome during early embryogenesis (BEAN *et al*. 2004). And in contrast to mammals and plants where DNA methylation is generally associated with gene inactivation, the inactive heterochromatinized paternal genome is *hypo*methylated compared to the active *hyper*methylated maternal genome (BONGIORNI *et al*. 1999; BONGIORNI and PRANTERA

2003). Hence, although DNA methylation appears to be widely associated with epigenetic phenomena, its true mechanistic role remains largely unknown.

Marking DNA with a methyl group appears to be an especially useful evolutionary invention that allows regulatory pathways to read additional information beyond the primary DNA sequence. For example, the simple addition of methyl groups to adenosine residues allows prokaryotes to distinguish not only self from non-self DNA, but also temporal information about self DNA (newly replicated DNA is not yet methylated) (BARRAS and MARINUS 1989; LOBNER-OLESEN *et al*. 2005). In epigenetic processes, DNA methylation, in conjunction with histone modification, give information about the parent-of-origin of a DNA sequence (as in imprinting) or the transcriptional competency of a particular locus (i.e. X inactivation and paramutation). It is not surprising, then, that DNA methylation (whether cytosine or adenine methylation) is one of the ubiquitous themes in biology used by diverse taxa for diverse biological processes.

Cells gain another level of information by adding histone modification to primary sequence and DNA methylation. It is now well-established that histone modifications play crucial roles in gene regulation. Histone modification allows cells to regulate the temporal and spatial expression of genes reversibly. This is especially relevant to epigenetic processes such as imprinting, which requires the same sequence of DNA to carry different states of information depending upon its lineage, and which requires resetting and establishing a new state to the DNA sequence.

It is now known that DMRs are not only sites of differential DNA modification, they are also sites of differential chromatin modification. The DMR of *Igf2*/*H19* (KHOSLA *et al*. 1999), *IC2*/*Kcnq1* (KANDURI *et al*. 2002), *PWS-IC* (SCHWEIZER *et al*. 1999), have

been shown to be sensitive to DNase I digestion in a parent-of-origin manner. Both activating (i.e. H3 and H4 acetylation) and de-activating histone modifications (H3 lysine 9 and lysine 27 methylation) have been shown (LEWIS and REIK 2006) at the DMRs. The pattern of histone modification appears to be correlated with the pattern of DNA methylation. For example, activating histone marks are found at unmethylated ICRs; whereas de-activating marks are generally associated with methylated ICRs. However, there are examples that histone modification at ICRs can occur independently of DNA methylation (LEWIS *et al*. 2004). Differential histone modification can also occur outside of ICRs to modulate imprinted gene expression. For example, at the *Igf2r* locus, it is allele-specific histone modification at the promoter, possibly in conjunction with DNA methylation at the ICR, that regulates imprinting (VU *et al*. 2004).

Involvement of differential histone modifications in imprinted loci is exemplified by certain insect species. In mealybugs, differential histone modifications mark the entire paternal genome as distinct from the maternal genome (BONGIORNI *et al*. 2001; COWELL *et al*. 2002; FERRARO *et al*. 2001). The paternal chromosomes, carrying de-activating histone marks, are eliminated in embryos that develop into males (BONGIORNI and PRANTERA 2003; BROWN 1959; GODAY and ESTEBAN 2001; KHOSLA *et al*. 2006). Similar mechanisms in Sciarid flies may mark selected paternal chromosomes for elimination (CROUSE 1960; GODAY and RUIZ 2002) during various stages of the organisms' life cycle.

Besides covalent histone modifications, one might imagine replacement of histone variants in modulating imprinting activities. Deposition of variant histones at a locus may give a different readout than from canonical histones (HAKE and ALLIS 2006). This would allow gene expression activities to be regulated at very defined regions along the

chromosome. Since deposition of variant histones can be replication-independent (HENIKOFF and AHMAD 2005), cells might achieve a more dynamic level of gene regulation that is independent of the cell cycle. As an example, local deposition of histone H3.3 could be expected to make the site transcriptionally active. Such a mechanism might be used, for example, to turn specific genes on or off in terminally differentiated cells.

Regulation of gene activities by variant histone replacement appears to be a ubiquitous process found in many systems and processes. Some examples include macroH2A and X chromosome inactivation in mouse (CHADWICK *et al*. 2001; CHADWICK and WILLARD 2003; CHANGOLKAR and PEHRSON 2006; HERNANDEZ-MUNOZ *et al*. 2005a), involvement of H1.1 in chromatin silencing and germline development in *C. elegans* (JEDRUSIK and SCHULZE 2001), AtMGH3 and sperm-specific chromatin remodeling in *Arabidopsis* (OKADA *et al*. 2005), His1-3 and drought stress in *Arabidopsis* (ASCENZI and GANTT 1997), H3.3 involvement in *Drosophila* sperm chromatin assembly (LOPPIN *et al*. 2005), H2A.Z and chromosome segregation in *Xenopus* (RIDGWAY *et al*. 2004), and H2A1.2 in mammalian genome stability (BASSING *et al*. 2002). Adding to the repertoire of biological processes that can be mediated by variant histone deposition, allele-specific variant histone macroH2A1 replacement has been found to occur at the ICRs of several known imprinted genes, including *Peg3*, *Igf2*/*H19*, *Gtl2*/*Dlk1*, and *Gnas* (CHOO *et al*. 2006).

non-coding RNAs: A characteristic feature of imprinted clusters is the presence of one or more non-coding transcript(s), often transcribed from the opposite direction as protein-coding genes in the cluster (ROYO *et al*. 2006). In mouse, the *Air* non-coding

RNA is transcribed from the paternal allele in an opposite direction as the imprinted *Igf2r* gene (LYLE *et al*. 2000). The promoter driving *Air* transcription is located within an intron of *Igf2r* (LYLE *et al*. 2000). Sleutels *et al*. has shown that the *Air* RNA has a regulatory role in the imprinting of flanking genes (SLEUTELS *et al*. 2002). Like *Air*, the paternally expressed *Kcnq1ot1* antisense RNA has been suggested to have a similar regulatory function at the *Ascl2-Cdkn1c-Kcnq1* imprinted cluster (MANCINI-DINARDO *et al*. 2006). These and other examples of regulation of an imprinted cluster by a non-coding antisense RNA is reminiscent of X chromosome inactivation in mammals.

Research into the RNAi and microRNA pathways in the last approximately eight years has led to the discovery of a vast number of biological processes regulated by small regulatory RNAs. It is not surprising, then, to find the involvement of these classes of non-coding RNAs in imprinting processes. Two loci that provide a good paradigm of involvement of small RNAs in imprinting are the Prader-Willi locus and the *Dlk1-Gtl2* (*callipyge* in sheep) clusters in humans (ROYO *et al*. 2006). We limit our discussion to the latter. The *Dlk1-Gtl2* locus contains a complex set of gene arrangement and expression patterns. In this gene cluster, the paternal chromosome expresses three protein-coding genes *Dlk1*, *Rtl1*/*Peg11*, and *Dio3* (GEORGES *et al*. 2003). In contrast, the maternal chromosome expresses various classes of non-coding RNAs including miRNAs, snoRNAs, and a transcript of the *Gtl2* gene. The maternally expressed *Peg11* non-coding RNA is transcribed in an anti-sense orientation to the paternal *Peg11*. The maternal *Peg11* transcript contains a cluster of miRNAs that mediate degradation of the paternal *Peg11* sense transcript via the RNAi pathway (DAVIS *et al*. 2005). Interestingly, *Dlk1*

(Delta-like 1) and *Rtl1* (Retrotransposon-like 1) appear to be remnants of retrotransposable elements (LYNCH and TRISTEM 2003; YOUNGSON *et al*. 2005).

Polycomb and Trithorax Group proteins: Polycomb group (PcG) and trithorax group (TrxG) proteins establish mitotically stable states. These proteins assemble into large complexes that recognize and bind target sequences called polycomb response elements (PREs) or trithorax response elements (TREs), respectively, to keep the target gene in the silenced (PcG-mediated) or active (TrxG-mediated) state through subsequent mitotic divisions (CERNILOGAR and ORLANDO 2005b). It should be noted that PREs and TREs are defined functionally: they are "DNA sequences to which PcG and TrxG complexes bind, directly or indirectly" (KLYMENKO *et al*. 2006). Thus, PcG/TrxG proteins are certainly capable of using affiliated proteins for DNA recognition and/or binding activities.

PcG and TrxG proteins mediate gene activities through chromatin remodeling, either by directly carrying out histone modifications or by recruiting histone modification factors and/or recruiting DNA modification factors to local sites (CAO *et al*. 2005; KUZMICHEV *et al*. 2004; KUZMICHEV *et al*. 2005; RINGROSE *et al*. 2004; VIRE *et al*. 2006; WANG *et al*. 2004). The simplest mechanism one can envision for PcG/TrxG-mediated epigenetic states is physical exclusion of *trans*-acting factors to target sequences, but there is little experimental support for this model (PIRROTTA 1997). Instead, experimental evidences from multiple labs seem to suggest that the mechanism is more complex, likely involving a series of recruitment and modification steps. Recognition and binding to PREs (assisted perhaps by specific histone modifications) is likely to be the initiating step that leads to recruitment of histone and/or DNA modification factors to the region, which

in turn leads to recruitment of additional factors that recognize specific histone or DNA methylation patterns (BANTIGNIES and CAVALLI 2006; CERNILOGAR and ORLANDO 2005a; HERNANDEZ-MUNOZ *et al*. 2005b; ZHANG *et al*. 2004). Once thought to maintain the repressed or active state of a target gene for the life of the organism, emerging evidences indicate that PcG/TrxG-mediated gene expression can be dynamic, allowing cells to switch fates during development (CHEN *et al*. 2005; FICZ *et al*. 2005; KLEBES *et al*. 2005; LEE *et al*. 2005).

PcG and TrxG proteins are found in many organisms and are involved in numerous developmental processes where maintenance of long term gene expression states is required. Not surprisingly, PcG/TrxG proteins have been found to be involved in genomic imprinting. The imprinted *MEDEA* gene in *Arabidopsis* is itself a PcG member (KINOSHITA *et al*. 1999). In mammals, PcG proteins have been shown to be required for parent-specific allele repression (LEWIS *et al*. 2004; MAGER *et al*. 2003; UMLAUF *et al*. 2004). Recently, the mammalian PcG genes *Eed* and *Ezh2* and a *Drosophila* PcG gene have been linked to DNA methylation (FERRES-MARCO *et al*. 2006; MAGER *et al*. 2003; VIRE *et al*. 2006). These are important findings that add to the repertoire of mechanisms by which PcG/TrxG proteins mediate genetic states.

The imprint life cycle

In imprinted genes, the same DNA sequence must be cycled between the male and female germline in different generations. This entails three distinct phases in the life cycle of an imprint: establishment of parent-specific imprints in the germline, maintenance and readout of the imprint in somatic cells of the progeny generation, and resetting and re-establishment of new imprints in the progeny germline (Figure 1.2). In

discussing an imprint life cycle, it is useful to keep the soma and germline distinct: imprints are maintained and read in somatic cells, but erasure (of the previous generation's imprints) and establishment of new imprints occur in the germline. How these processes are thought to occur is discussed below.

Establishment of the imprinted state: Early studies of imprinting in mice indicated germline passage as a requirement for the establishment of gamete-specific imprints (TUCKER *et al*. 1996). Thus, imprints are established during oogenesis and gametogenesis. Since the same DNA sequence must alternate between male and female gametes, imprints are likely to involve gamete-specific factors that add gamete-specific marks. Cells employ multiple chromatin marks including DNA methylation, histone modification, and variant histone deposition.

Gamete-specific DNA methylation patterns have been known since the earliest days of studies on genomic imprinting (CHAILLET *et al*. 1991; REIK *et al*. 1987; SWAIN *et al*. 1987). Differential methylation is often targeted to DMRs. DMRs are divided into two classes based on the timing of their acquisition of methylation. Primary DMRs acquire gamete-specific methylation during gametogenesis and thus carry information about gamete origin. Primary DMRs are the targets of *de novo* DNA methyltransferases such as DNMT3a and DNMT3L (HATA *et al*. 2002; KANEDA *et al*. 2004). Secondary DMRs acquire gamete-specific methylation after fertilization and may carry information about tissue specificity, for example. In mammals, secondary DMRs are targeted by the maintenance DNA methyltransferase DNMT1 (HOWELL *et al*. 2001).

DMRs are also sites of gamete-specific histone modifications. Both activating and de-activating modifications have been found that correlate with the parent-specific

expression of the locus. *Snrpn*, *Igf2r*, and *U2af1-rs1* are all expressed from the paternally allele (FOURNIER *et al*. 2002). Analysis of the DMRs of these imprinted loci revealed that the paternal DMRs contain activating histone modifications (i.e. H3-K4 methylation and H3-K9/K14 acetylation); while the corresponding maternal DRM's contain deactivating histone marks (i.e. H3-K9 methylation) (FOURNIER *et al*. 2002). The maternal *U2af1-rs1* DMR is hypermethylated and is associated with the MBD proteins MeCP2, MBD1, and MBD2 (FOURNIER *et al*. 2002). Thus, the deactivating histone marks on the maternal DMR is associated with heavy methylation, also a deactivating chromatin mark. Numerous examples of differential histone modifications at DMRs have been found for other imprinted clusters, including PWS/AS (XIN *et al*. 2001) and BWS (HIGASHIMOTO *et al*. 2003).

Deposition of variant histones during gametogenesis has been documented in a number of species (AKHMANOVA *et al*. 1997; HENNIG 2003; LOPEZ-ALANON *et al*. 1997; MARTIANOV *et al*. 2005; NICKEL *et al*. 1987; WATSON *et al*. 1999). Variant histone replacement during gametogenesis is believed to help with chromatin packaging, particularly in sperm. Recently, Choo *et al*. reported the allele-specific deposition of histone variant macroH2A1 at some imprinted loci they investigated (CHOO *et al*. 2006). Although the study does not address the question of whether deposition occurred during gametogenesis or post-fertilization (the analysis was done using mouse somatic tissue), it does reaffirm the numerous strategies cells employ to modulate epigenetic states.

Polycomb group proteins are known to be required for germline development in multiple organisms examined (HOLDEMAN *et al*. 1998; JOHNSON *et al*. 2004; KELLY and FIRE 1998; KORF *et al*. 1998; LAWRENCE *et al*. 1983; PARO and ZINK 1992). PcG/TrxG

protein activities are detected during embryogenesis, during which they establish and maintain gene activity states. The function of PcG/TrxG proteins during embryogenesis is well-documented in *Drosophila* (CHANAS and MASCHAT 2005; GOULD *et al*. 1990; ORLANDO *et al*. 1998; RILEY *et al*. 1987), but has also been reported in mammals (FAUST *et al*. 1998; HOBERT *et al*. 1996; O'CARROLL *et al*. 2001; SHUMACHER *et al*. 1996; TAKIHARA *et al*. 1997; YU *et al*. 1998), plants (GOODRICH *et al*. 1997; GROSSNIKLAUS *et al*. 1998; GUITTON and BERGER 2005), and possibly in *C. elegans* as well (CHAMBERLIN and THOMAS 2000; ZHANG *et al*. 2003). However, as far as we know, establishment of epigenetic states by PcG proteins during gametogenesis has not been reported. PcG proteins have been shown to exist during oogenesis in *Drosophila* but has not been shown to localize to the oocyte nucleus (PARO and ZINK 1992). This finding excludes any role for PcG in the establishment of epigenetic states during oogenesis.

Current available data appears to suggest that PcG/TrxG proteins have a maintenance role, rather than an establishment role, in regulating gene activities. In principle, cells should be able to set epigenetic states during gametogenesis using PcG/Trx proteins. Two mechanisms can be envisioned. These two modes of action of PcG/TrxG proteins have been reported in regulating gene expression post-fertilization.

Mechanism 1: PcG/TrxG proteins could function as recruiters that attract DNA and histone modification factors (histone acetyltransferases, histone de-acetylases, histone sumoylation/phosphorylation/ubiquitination factors) to target sites. In this scenario, the PcG/TrxG themselves possess histone modification activities, laying down histone marks at target PREs. By reading this PcG/TrxG-mediated histone code, other histone modification factors would be recruited to the site. Recently,

the human homolog of PRC1(hPRC1L) has been demonstrated to contain the ubiquitinase activity for histone H2A (WANG *et al*. 2004). The downstream event has been demonstrated to be initiation of X inactivation (FANG *et al*. 2004).

ESC/E(Z), in conjunction with PRC1, constitute the two polycomb complexes found in *Drosophila* (REFERENCE). Two groups have simultaneously demonstrated that ESC/E(Z) contains H3-K9/K27 methyltransferase activity (CZERMIN *et al*. 2002; MULLER *et al*. 2002). The downstream event(s) is not known, but H3-K9/K27 dimethylation could potentially serve as a code for other chromatin proteins.

Mechanism 2: PcG/TrxG proteins could be recruited to target sites by reading the histone code laid down by other chromatin modification factors. In this mode of action, PcG/TrxG proteins could function as bridges between the various effectors of chromatin remodeling. For example, Pc is a subunit of the *Drosophila* PRC1 polycomb complex that recognizes methylated H3-K27 (ZHANG *et al*. 2004). When present in the *Ubx* PRE, H3-K27-methyl acts as a code to attract Pc, which in turn recruits the PRC1 complex to the *Ubx* PRE to repress *Ubx* expression (ZHANG *et al*. 2004). A similar mechanism has been found to operate at the *Drosophila* PcG *Pho* locus (KLYMENKO *et al*. 2006).

Maintenance and readout of the imprinted state: In normal fertilization events in mammals, both the paternal and maternal genomes of somatic cells undergo genome-wide de-methylation shortly after fertilization (BARTON *et al*. 2001; MAYER *et al*. 2000; OSWALD *et al*. 2000; SANTOS *et al*. 2002). An interesting aspect of imprinted genes is their resistance to this genome-wide de-methylation event. What are the mechanisms of

maintenance? In mammals, both DNA methylation and histone modification marks are maintained. DNA methylation patterns are maintained by the maintenance DNA methyltransferase *Dnmt1* (HOWELL *et al*. 2001). For example, the *H19* gene is normally only expressed from the maternal allele, but in a *Dnmt1* loss-of-function background, the paternal allele is also expressed (LI *et al*. 1993). How exactly the methyltransferase knows which sequence to target is not known, but evidence suggests that primary DMRs may serve as guides.

Somatic cells must not only maintain the parental generation's imprints, but must read the imprints and express the proper alleles. What is the mechanism(s) by which this occurs? Not all cells will read and respond to an imprint in the same manner. There must be tissue-specific regulatory elements that regulate the spatial and temporal expression of imprints. For example, the imprinted *Igf2* gene is expressed from the paternal allele in most tissues. In the liver, alternative promoter usage leads to biallelic *Igf2* expression (VU and HOFFMAN 1994). Hagège *et al*. has shown that additional tissue-specificity of *Igf2* requires a *cis*-acting sequence located 3' of the *H19* gene (HAGEGE *et al*. 2006). Further studies by Ohno *et al*. indicated that there is also temporal regulation of the *Igf2* imprint, transitioning from biallelic to gamete-origin-specific transcription during the blastocyst stage (OHNO *et al*. 2001). Numerous other studies by different groups confirmed additional imprinted loci being imprinted in a spatial and/or temporal manner, including *Neurabin*, *Pon2*, and *Pon3* (ONO *et al*. 2003), *Tnfrh1* (CLARK *et al*. 2002), *Dio3* (TSAI *et al*. 2002), and *Nesp* and *Nespas* (BALL *et al*. 2001). Thus, establishment of the imprint simply marks the two parental alleles as different. It is up to the next generation to maintain, interpret, and express the imprints in the proper spatial and temporal manner.

Imprint erasure and resetting: While the somatic cells must maintain, read, and express the parental imprints, the germline has the task of resetting the previous generation's imprints and laying down its own sex-specific imprints. In mammals at least, parental imprints are reset during embryonic development (REIK and WALTER 2001). During early embryonic development, the primordial germ cells undergo global demethylation (i.e. erasure). Two general strategies have been adopted to determine the exact timing of this event. One class of experiments aimed at determining methylation status of specific imprinted loci in variously staged gametes; another class of experiments employed nuclear transfer or cloning of mice from primordial germ cells, the idea being that only properly imprinted nuclei are competent to direct normal development of embryos. Collectively, these two types of experiments carried out by numerous groups have led to three general conclusions about imprint erasure in the germline: (1) erasure occurs early, around 9.5-10.5 days post coitum; (2) erasure is asynchronous, that is, it does not occur simultaneously for all imprinted genes; (3) erasure is rapid, possibly involving an active mechanism of de-methylation (DURCOVA-HILLS *et al*. 2001; HAJKOVA *et al*. 2002; LEE *et al*. 2002; LUCIFERO *et al*. 2004; SATO *et al*. 2003; YAMAZAKI *et al*. 2005).

Two models were originally put forth to account for the mechanism of imprint erasure (ROSSANT 1993). One postulates that the imprints of the same-sex parent are preserved, while the imprints of the opposite-sex parent are reversed The other model posits that both imprints are erased, followed by re-establishment of gamete-specific imprints. A preponderance of data indicate the second model to be more likely.

Selective inactivation versus selective activation of imprinted genes

The steps involved in imprinted gene expression are: (1) mark the two parental alleles as different; (2) maintain and read the imprinted marks in somatic cells; (3) reset the parental marks and re-establish gamete-specific marks (during gametogenesis). In mammals, these three steps lead to selective inactivation of the non-expressed allele, and the cycling of a DNA sequence between the active or inactive state depending on its gamete of origin. Remarkably, the imprint life cycle operates on the entire paternal genome in the mealybugs (BONGIORNI *et al.* 1999; BONGIORNI *et al.* 2001; BONGIORNI and PRANTERA 2003; KHOSLA *et al.* 2006). In the sciara flies, various paternal chromosomes are selectively inactivated and eliminated during various phases of the life cycle (BROWN and CHANDRA 1977; CROUSE 1960; GODAY and ESTEBAN 2001).

Plants exhibit very divergent biology compared to animals. It is not surprising, then, that plants have evolved a different strategy to imprint genes. In plants, it appears that the expressed allele of an imprinted gene is selectively activated (SCOTT and SPIELMAN 2004; SCOTT and SPIELMAN 2006). *MEDEA* and *FWA* are two well-characterized imprinted genes in *Arabidopsis*, both expressed only from the female gamete (CHAUDHURY *et al.* 1997; GROSSNIKLAUS *et al.* 1998; KINOSHITA *et al.* 1999; KINOSHITA *et al.* 2004; KIYOSUE *et al.* 1999). The default state of both genes is inactive, and is associated with heavy DNA methylation by the maintenance methyltransferase *MET1*, a homolog of the human maintenance methyltransferase *Dnmt1* (KANKEL *et al.* 2003; RONEMUS *et al.* 1996; XIAO *et al.* 2003). Activation of *MEDEA* and *FWA* in female gametes requires *DEMETER*, a DNA glycosylase implicated to have DNA de-methylation activity (CHOI *et al.* 2002; KINOSHITA *et al.* 2004; XIAO *et al.* 2003).

Hence, different organisms employ different strategies to achieve the same end result: selective expression of only one allele of imprinted genes. Whatever route is employed, the end result is differential chromatin states between the two parental alleles. *Trans*-acting factors then read the information carried in the chromatin and affect spatial and/or temporal gene expression.

C. elegans as a model organism to study epigenetic gene regulation

Caenorhabditis elegans is a tiny non-parasitic roundworm originally chosen by Sidney Brenner in the 1970's for its many attributes as an ideal model organism (BRENNER 1974). Among the characteristics that make this organism a good system in which to study genetics are (1) small size (allowing for cultivation of vast numbers using relatively small amounts of space and resources), (2) short life cycle, (3) large brood sizes, (4) transparency (easy for microscopy), (5) a relatively small genome (100 megabases packaged into five autosomes and one X chromosome), (6) relatively simple to cultivate in laboratory settings, and (7) can be kept indefinitely in liquid nitrogen.

In the past three decades, scientists using *C. elegans* as a model system have made many seminal contributions to the field of biology. The first complete lineage of a eukaryote was that of *C. elegans* (SULSTON *et al.* 1983). Knowledge of cell lineages opened up avenues for studies in other biological processes, such as apoptosis and organogenesis. Pioneering work by Horvitz, Hedgecock, and others led to one of the very first elucidations of the apoptotic pathway (ELLIS and HORVITZ 1986; HEDGECOCK *et al.* 1983). The first molecular genetic study of heterochrony was in *C. elegans* (AMBROS and HORVITZ 1984). The importance of this work laid dormant for another decade, when elucidation of the mechanism of the RNAi pathway led to the discovery of microRNAs

and siRNAs as related gene regulatory pathways (FIRE *et al*. 1998; GRISHOK *et al*. 2000; GRISHOK *et al*. 2001; LAGOS-QUINTANA *et al*. 2001; LAU *et al*. 2001; LEE and AMBROS 2001; PARRISH *et al*. 2000; PARRISH and FIRE 2001; SIJEN *et al*. 2001). So pervasive are small regulatory RNAs in biological processes that there have been "heretical" suggestions that their place as regulatory molecules would likely be commensurate to that of proteins.

In 1998, the *C. elegans* genome became the first eukaryotic genome to be sequenced (CONSORTIUM 1998). As valuable as the data that could be mined from a complete genome sequence is the experience learned from such an endeavor. Lessons learned from the *C. elegans* genome sequence project improved the efficiency and speed of other genome sequencing projects that followed.

The work described in this dissertation has been an attempt to understand epigenetic regulation of gene expression, using *C. elegans* as a model. The successful use of transgenes to study gene expression has been a powerful tool employed in many fields of biology. In *C. elegans*, as in many organisms, introduction of a transgene into the organism leads to the formation of long, often highly repeated, transgene arrays. In *C. elegans*, as in many other organisms, highly repetitive transgene arrays tend to be silenced by the organism. While silencing of transgene arrays may be an annoyance to researchers whose intentions are to achieve high expression (i.e. gene therapy, transgenic rescue, ectopic gene expression and/or protein production), the mechanism of silencing in and of itself is a biological process worthy of investigation. We employed transgenic techniques in an attempt to elucidate silencing mechanisms. In particular, we were interested in identifying endogenous *C. elegans* factors that recognize and silence foreign

DNA. As is usually the case in scientific research, unexpected roadblocks along the way led to the observation of a previously uncharacterized process: that *C. elegans* can apparently imprint DNA. Although we do not know the mechanism by which *C. elegans* imprint DNA, lessons from other well-characterized systems point to mechanisms modulated by chromatin. To this end, we have also begun to develop tools to study *C. elegans* chromatin on a genome-wide scale.

FIGURES

Figure 1.1. DMR imprint control at the *Igf2/H19* locus. The DMR controlling allele-specific expression at the *Igf2/H19* imprinted cluster is represented by the yellow bar. It is distinct from DMR1, another *cis*-acting element at the *Igf2/H19* locus. When bound by CTCF, the DMR acts as an insulator and blocks activation of the maternal *Igf2* promoter by the downstream enhancer. Black lollipops represent methylated cytosine residues. Black arrows indicate transcription. [enh = enhancer]

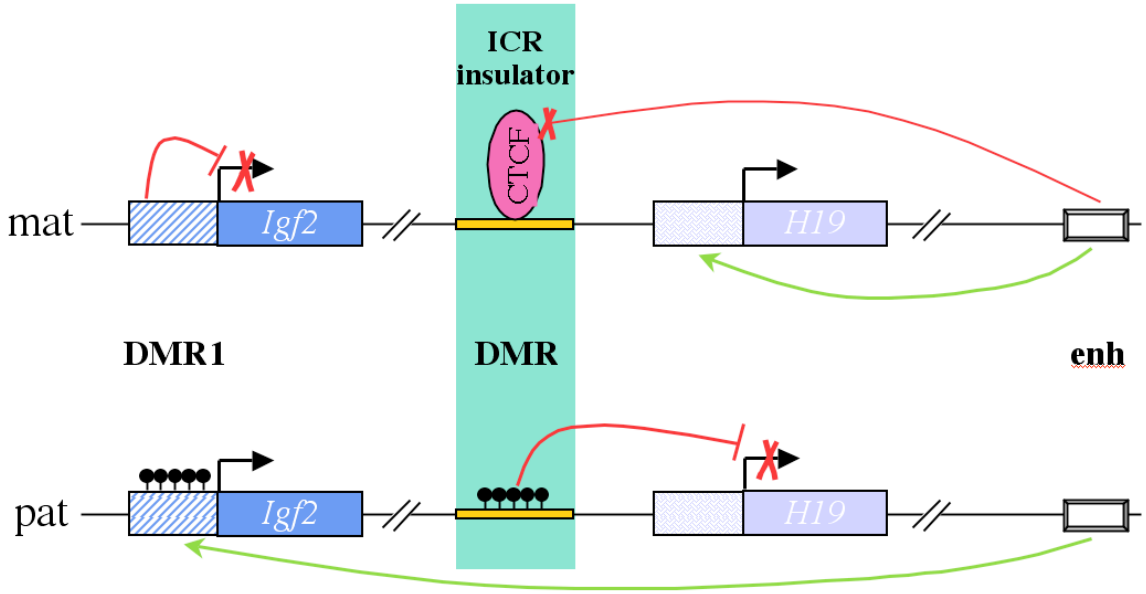
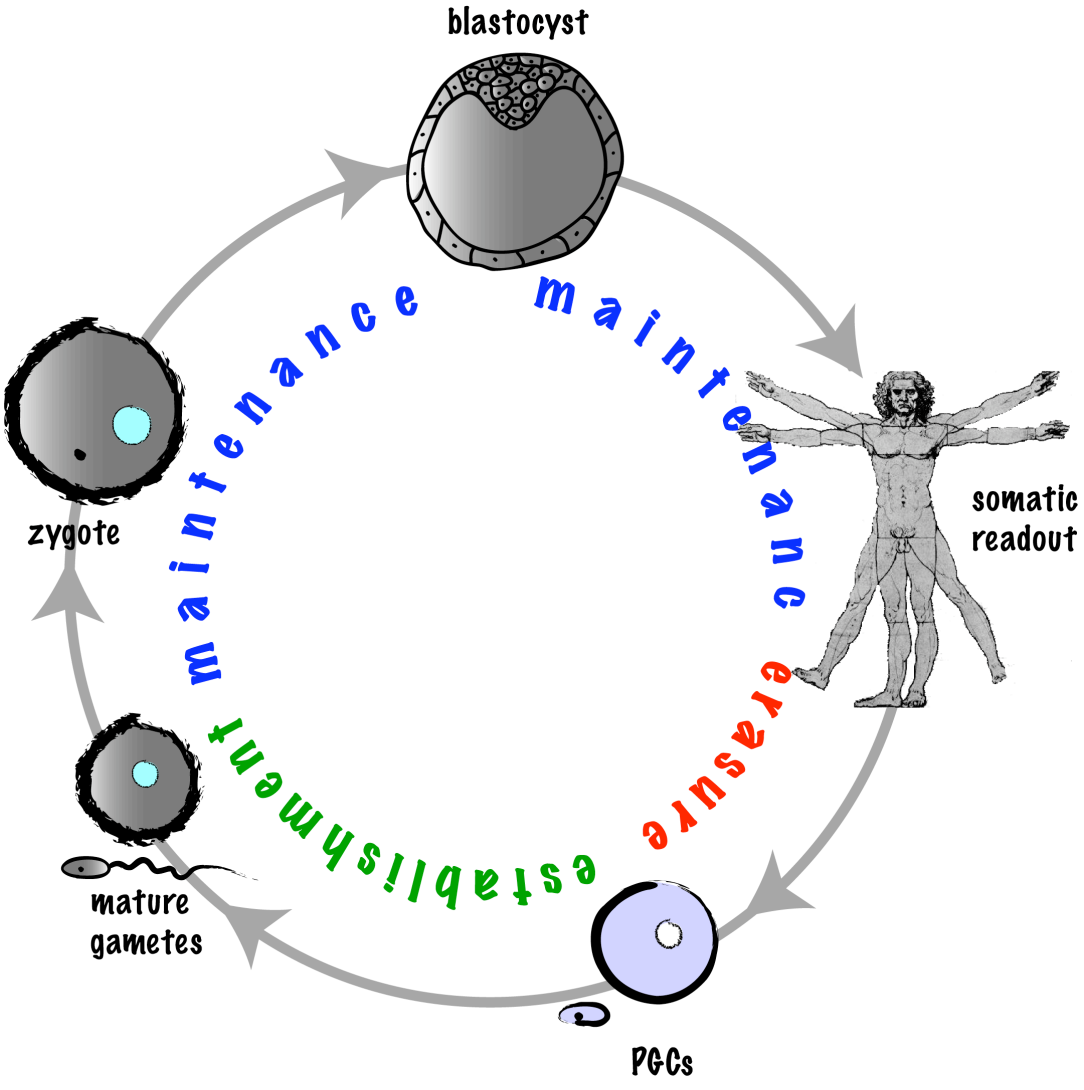


Figure 1.2. The imprint life cycle. Beginning at fertilization, the parental imprints are maintained in the blastocyst and all subsequent somatic lineages. In somatic cells of the embryo and adult organism, imprints are maintained and read. In the developing germline, parental imprints undergo erasure in primordial germ cells (PGCs). As germ cells mature, sex-specific imprints are re-established. Fertilization completes the imprint life cycle.



REFERENCES

AINSCOUGH, J. F., T. KOIDE, M. TADA, S. BARTON and M. A. SURANI, 1997 Imprinting of *Igf2* and *H19* from a 130 kb YAC transgene. Development **124:** 3621-3632.

AKHMANOVA, A., K. MIEDEMA, Y. WANG, M. VAN BRUGGEN, J. H. BERDEN *et al.*, 1997 The localization of histone H3.3 in germ line chromatin of Drosophila males as established with a histone H3.3-specific antiserum. Chromosoma **106:** 335-347.

AMBROS, V., and H. R. HORVITZ, 1984 Heterochronic mutants of the nematode Caenorhabditis elegans. Science **226:** 409-416.

ASCENZI, R., and J. S. GANTT, 1997 A drought-stress-inducible histone gene in Arabidopsis thaliana is a member of a distinct class of plant linker histone variants. Plant Mol Biol **34:** 629-641.

BALL, S. T., C. M. WILLIAMSON, C. HAYES, T. HACKER and J. PETERS, 2001 The spatial and temporal expression pattern of Nesp and its antisense Nespas, in mid-gestation mouse embryos. Mech Dev **100:** 79-81.

BANTIGNIES, F., and G. CAVALLI, 2006 Cellular memory and dynamic regulation of polycomb group proteins. Curr Opin Cell Biol **18:** 275-283.

BARRAS, F., and M. G. MARINUS, 1989 The great GATC: DNA methylation in *E. coli*. Trends Genet **5:** 139-143.

BARTON, S. C., K. L. ARNEY, W. SHI, A. NIVELEAU, R. FUNDELE *et al.*, 2001 Genome-wide methylation patterns in normal and uniparental early mouse embryos. Hum Mol Genet **10:** 2983-2987.

BASSING, C. H., K. F. CHUA, J. SEKIGUCHI, H. SUH, S. R. WHITLOW *et al.*, 2002 Increased ionizing radiation sensitivity and genomic instability in the absence of histone H2AX. Proc Natl Acad Sci U S A **99:** 8173-8178.

BAYLIN, S. B., and J. E. OHM, 2006 Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction? Nat Rev Cancer **6:** 107-116.

BEAN, C. J., C. E. SCHANER and W. G. KELLY, 2004 Meiotic pairing and imprinted X chromatin assembly in *Caenorhabditis elegans*. Nat Genet **36:** 100-105.

BEATTY, L., R. WEKSBERG and P. D. SADOWSKI, 2006 Detailed analysis of the methylation patterns of the KvDMR1 imprinting control region of human chromosome 11. Genomics **87:** 46-56.

BELL, A. C., and G. FELSENFELD, 2000 Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature **405:** 482-485.

BONGIORNI, S., O. CINTIO and G. PRANTERA, 1999 The relationship between DNA methylation and chromosome imprinting in the coccid Planococcus citri. Genetics **151:** 1471-1478.

BONGIORNI, S., M. MAZZUOLI, S. MASCI and G. PRANTERA, 2001 Facultative heterochromatization in parahaploid male mealybugs: involvement of a heterochromatin-associated protein. Development **128:** 3809-3817.

BONGIORNI, S., and G. PRANTERA, 2003 Imprinted facultative heterochromatization in mealybugs. Genetica **117:** 271-279.

BOURC'HIS, D., G. L. XU, C. S. LIN, B. BOLLMAN and T. H. BESTOR, 2001 Dnmt3L and the establishment of maternal genomic imprints. Science **294:** 2536-2539.

BRENNER, S., 1974 The genetics of *Caenorhabditis elegans*. Genetics **77:** 71-94.

BROWN, S. W., 1959 Lecanoid chromosome behavior in three more families of the Coccoidea (Homoptera). Chromosoma **10:** 278-300.

BROWN, S. W., and H. S. CHANDRA, 1977 *Chromosome imprinting and the differential regulation of homologous chromosomes.* Academic Press, New York.

CAO, R., Y. TSUKADA and Y. ZHANG, 2005 Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. Mol Cell **20:** 845-854.

CERNILOGAR, F. M., and V. ORLANDO, 2005a Epigenome programming by Polycomb and Trithorax proteins. Biochem Cell Biol **83:** 322-331.

CERNILOGAR, F. M., and V. ORLANDO, 2005b Epigenome programming by Polycomb and Trithorax proteins. Biochem. Cell Biol **83:** 322-331.

CHADWICK, B. P., C. M. VALLEY and H. F. WILLARD, 2001 Histone variant macroH2A contains two distinct macrochromatin domains capable of directing macroH2A to the inactive X chromosome. Nucleic Acids Res **29:** 2699-2705.

CHADWICK, B. P., and H. F. WILLARD, 2003 Chromatin of the Barr body: histone and non-histone proteins associated with or excluded from the inactive X chromosome. Hum Mol Genet **12:** 2167-2178.

CHAILLET, J. R., T. F. VOGT, D. R. BEIER and P. LEDER, 1991 Parental-specific methylation of an imprinted transgene is established during gametogenesis and progressively changes during embryogenesis. Cell **66:** 77-83.

CHAMBERLIN, H. M., and J. H. THOMAS, 2000 The bromodomain protein LIN-49 and trithorax-related protein LIN-59 affect development and gene expression in *Caenorhabditis elegans*. Development **127:** 713-723.

CHANAS, G., and F. MASCHAT, 2005 Tissue specificity of hedgehog repression by the Polycomb group during Drosophila melanogaster development. Mech Dev **122:** 975-987.

CHANGOLKAR, L. N., and J. R. PEHRSON, 2006 macroH2A1 Histone Variants Are Depleted on Active Genes but Concentrated on the Inactive X Chromosome. Mol Cell Biol **26:** 4410-4420.

CHAUDHURY, A. M., L. MING, C. MILLER, S. CRAIG, E. S. DENNIS *et al.*, 1997 Fertilization-independent seed development in Arabidopsis thaliana. Proc Natl Acad Sci U S A **94:** 4223-4228.

CHEN, X., M. HILLER, Y. SANCAK and M. T. FULLER, 2005 Tissue-specific TAFs counteract Polycomb to turn on terminal differentiation. Science **310:** 869-872.

CHOI, Y., M. GEHRING, L. JOHNSON, M. HANNON, J. J. HARADA *et al.*, 2002 DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis. Cell **110:** 33-42.

CHOO, J. H., J. D. KIM, J. H. CHUNG, L. STUBBS and J. KIM, 2006 Allele-specific deposition of macroH2A1 in imprinting control regions. Hum Mol Genet **15:** 717-724.

CLARK, L., M. WEI, G. CATTORETTI, C. MENDELSOHN and B. TYCKO, 2002 The Tnfrh1 (Tnfrsf23) gene is weakly imprinted in several organs and expressed at the trophoblast-decidua interface. BMC Genet **3:** 11.

CONSORTIUM, 1998 Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science **282:** 2012-2018.

COWELL, I. G., R. AUCOTT, S. K. MAHADEVAIAH, P. S. BURGOYNE, N. HUSKISSON *et al.*, 2002 Heterochromatin, HP1 and methylation at lysine 9 of histone H3 in animals. Chromosoma **111:** 22-36.

CROUSE, H. V., 1960 The controlling element in sex chromosome behavior in Sciara. Genetics **45:** 1429-1443.

CUBAS, P., C. VINCENT and E. COEN, 1999 An epigenetic mutation responsible for natural variation in floral symmetry. Nature **401:** 157-161.

CZERMIN, B., R. MELFI, D. MCCABE, V. SEITZ, A. IMHOF *et al.*, 2002 Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell **111:** 185-196.

DAVIS, E., F. CAIMENT, X. TORDOIR, J. CAVAILLE, A. FERGUSON-SMITH *et al.*, 2005 RNAi-mediated allelic trans-interaction at the imprinted Rtl1/Peg11 locus. Curr Biol **15:** 743-749.

DURCOVA-HILLS, G., J. AINSCOUGH and A. MCLAREN, 2001 Pluripotential stem cells derived from migrating primordial germ cells. Differentiation **68:** 220-226.

EASTMAN, E. M., R. M. GOODMAN, B. F. ERLANGER and O. J. MILLER, 1980 5-methylcytosine in the DNA of the polytene chromosomes of the diptera Sciara coprophila, Drosophila melanogaster and D. persimilis. Chromosoma **79:** 225-239.

ELLIS, H. M., and H. R. HORVITZ, 1986 Genetic control of programmed cell death in the nematode C. elegans. Cell **44:** 817-829.

FANG, J., T. CHEN, B. CHADWICK, E. LI and Y. ZHANG, 2004 Ring1b-mediated H2A ubiquitination associates with inactive X chromosomes and is involved in initiation of X inactivation. J Biol Chem **279:** 52812-52815.

FAUST, C., K. A. LAWSON, N. J. SCHORK, B. THIEL and T. MAGNUSON, 1998 The Polycomb-group gene eed is required for normal morphogenetic movements during gastrulation in the mouse embryo. Development **125:** 4495-4506.

FERRARO, M., G. L. BUGLIA and F. ROMANO, 2001 Involvement of histone H4 acetylation in the epigenetic inheritance of different activity states of maternally and paternally derived genomes in the mealybug Planococcus citri. Chromosoma **110:** 93-101.

FERRES-MARCO, D., I. GUTIERREZ-GARCIA, D. M. VALLEJO, J. BOLIVAR, F. J. GUTIERREZ-AVINO *et al.*, 2006 Epigenetic silencers and notch collaborate to promote malignant tumours by Rb silencing. Nature **439:** 430-436.

FICZ, G., R. HEINTZMANN and D. J. ARNDT-JOVIN, 2005 Polycomb group protein complexes exchange rapidly in living Drosophila. Development **132:** 3963-3976.

FIRE, A., S. XU, M. K. MONTGOMERY, S. A. KOSTAS, S. E. DRIVER *et al.*, 1998 Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature **391:** 806-811.

FITZPATRICK, G. V., P. D. SOLOWAY and M. J. HIGGINS, 2002 Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. Nat Genet **32:** 426-431.

FOURNIER, C., Y. GOTO, E. BALLESTAR, K. DELAVAL, A. M. HEVER *et al.*, 2002 Allele-specific histone lysine methylation marks regulatory regions at imprinted mouse genes. Embo J **21:** 6560-6570.

GEHRING, M., J. H. HUH, T. F. HSIEH, J. PENTERMAN, Y. CHOI *et al.*, 2006 DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. Cell **124:** 495-506.

GEORGES, M., C. CHARLIER and N. COCKETT, 2003 The callipyge locus: evidence for the trans interaction of reciprocally imprinted genes. Trends Genet **19:** 248-252.

GODAY, C., and M. R. ESTEBAN, 2001 Chromosome elimination in sciarid flies. Bioessays **23:** 242-250.

GODAY, C., and M. F. RUIZ, 2002 Differential acetylation of histones H3 and H4 in paternal and maternal germline chromosomes during development of sciarid flies. J Cell Sci **115:** 4765-4775.

GOLL, M. G., and T. H. BESTOR, 2005 Eukaryotic cytosine methyltransferases. Annu Rev Biochem **74:** 481-514.

GOODRICH, J., P. PUANGSOMLEE, M. MARTIN, D. LONG, E. M. MEYEROWITZ *et al.*, 1997 A Polycomb-group gene regulates homeotic gene expression in Arabidopsis. Nature **386:** 44-51.

GOULD, A. P., R. Y. LAI, M. J. GREEN and R. A. WHITE, 1990 Blocking cell division does not remove the requirement for Polycomb function in Drosophila embryogenesis. Development **110:** 1319-1325.

GRISHOK, A., H. TABARA and C. C. MELLO, 2000 Genetic requirements for inheritance of RNAi in C. elegans. Science **287:** 2494-2497.

GRISHOK, A., A. E. PASQUINELLI, D. CONTE, N. LI, S. PARRISH *et al*., 2001 Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. Cell **106:** 23-34.

GROSSNIKLAUS, U., J. P. VIELLE-CALZADA, M. A. HOEPPNER and W. B. GAGLIANO, 1998 Maternal control of embryogenesis by *MEDEA*, a polycomb group gene in Arabidopsis. Science **280:** 446-450.

GUITTON, A. E., and F. BERGER, 2005 Loss of function of MULTICOPY SUPPRESSOR OF IRA 1 produces nonviable parthenogenetic embryos in Arabidopsis. Curr Biol **15:** 750-754.

GUSTAFSSON, A., 1979 Linnaeus' Peloria: The History of a Monster. Theor. Appl. Genet. **54:** 241-248.

GUTIERREZ, A., and R. J. SOMMER, 2004 Evolution of *dnmt-2* and *mbd-2*-like genes in the free-living nematodes *Pristionchus pacificus*, *Caenorhabditis elegans* and *Caenorhabditis briggsae*. Nucleic Acids Res **32:** 6388-6396.

HAGEGE, H., R. NASSER, M. WEBER, L. MILLIGAN, N. APTEL *et al*., 2006 The 3' portion of the mouse H19 Imprinting-Control Region is required for proper tissue-specific expression of the Igf2 gene. Cytogenet Genome Res **113:** 230-237.

HAJKOVA, P., S. ERHARDT, N. LANE, T. HAAF, O. EL-MAARRI *et al*., 2002 Epigenetic reprogramming in mouse primordial germ cells. Mech Dev **117:** 15-23.

HAKE, S. B., and C. D. ALLIS, 2006 Histone H3 variants and their potential role in indexing mammalian genomes: The "H3 barcode hypothesis". Proc Natl Acad Sci U S A **103:** 6428-6435.

HARK, A. T., C. J. SCHOENHERR, D. J. KATZ, R. S. INGRAM, J. M. LEVORSE *et al.*, 2000 CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature **405:** 486-489.

HATA, K., M. OKANO, H. LEI and E. LI, 2002 Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development **129:** 1983-1993.

HEDGECOCK, E. M., J. E. SULSTON and J. N. THOMSON, 1983 Mutations affecting programmed cell deaths in the nematode Caenorhabditis elegans. Science **220:** 1277-1279.

HENIKOFF, S., and K. AHMAD, 2005 Assembly of variant histones into chromatin. Annu Rev Cell Dev Biol **21:** 133-153.

HENNIG, W., 2003 Chromosomal proteins in the spermatogenesis of Drosophila. Chromosoma **111:** 489-494.

HERNANDEZ-MUNOZ, I., A. H. LUND, P. VAN DER STOOP, E. BOUTSMA, I. MUIJRERS *et al.*, 2005a Stable X chromosome inactivation involves the PRC1 Polycomb complex and requires histone MACROH2A1 and the CULLIN3/SPOP ubiquitin E3 ligase. Proc Natl Acad Sci U S A **102:** 7635-7640.

HERNANDEZ-MUNOZ, I., P. TAGHAVI, C. KUIJL, J. NEEFJES and M. VAN LOHUIZEN, 2005b Association of BMI1 with polycomb bodies is dynamic and requires PRC2/EZH2 and the maintenance DNA methyltransferase DNMT1. Mol Cell Biol **25:** 11047-11058.

HIGASHIMOTO, K., T. URANO, K. SUGIURA, H. YATSUKI, K. JOH *et al.*, 2003 Loss of CpG methylation is strongly correlated with loss of histone H3 lysine 9 methylation at

DMR-LIT1 in patients with Beckwith-Wiedemann syndrome. Am J Hum Genet **73:** 948-956.

HOBERT, O., I. SURES, T. CIOSSEK, M. FUCHS and A. ULLRICH, 1996 Isolation and developmental expression analysis of Enx-1, a novel mouse Polycomb group gene. Mech Dev **55:** 171-184.

HODGKIN, J., 1994 Epigenetics and the maintenance of gene activity states in Caenorhabditis elegans. Dev Genet **15:** 471-477.

HOLDEMAN, R., S. NEHRT and S. STROME, 1998 MES-2, a maternal protein essential for viability of the germline in Caenorhabditis elegans, is homologous to a Drosophila Polycomb group protein. Development **125:** 2457-2467.

HOLMES, R., and P. D. SOLOWAY, 2006 Regulation of imprinted DNA methylation. Cytogenet Genome Res **113:** 122-129.

HORIKE, S., K. MITSUYA, M. MEGURO, N. KOTOBUKI, A. KASHIWAGI *et al.*, 2000 Targeted disruption of the human LIT1 locus defines a putative imprinting control element playing an essential role in Beckwith-Wiedemann syndrome. Hum Mol Genet **9:** 2075-2083.

HOWELL, C. Y., T. H. BESTOR, F. DING, K. E. LATHAM, C. MERTINEIT *et al.*, 2001 Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. Cell **104:** 829-838.

HURST, L. D., G. MCVEAN and T. MOORE, 1996 Imprinted genes have few and small introns. Nat Genet **12:** 234-237.

JEDRUSIK, M. A., and E. SCHULZE, 2001 A single histone H1 isoform (H1.1) is essential for chromatin silencing and germline development in Caenorhabditis elegans. Development **128:** 1069-1080.

JIANG, Y. H., J. BRESSLER and A. L. BEAUDET, 2004 Epigenetics and human disease. Annu Rev Genomics Hum Genet **5:** 479-510.

JOHNSON, J., J. CANNING TILLY and J. TILLY, 2004 Expression of Bmi-1, a polycomb group gene family member, in postnatal mouse ovary, pp. in *Society for the Study of Reproduction*, Vancouver, British Columbia, Canada.

JULLIEN, P. E., T. KINOSHITA, N. OHAD and F. BERGER, 2006 Maintenance of DNA Methylation during the Arabidopsis Life Cycle Is Essential for Parental Imprinting. Plant Cell **18:** 1360-1372.

KANDURI, C., V. PANT, D. LOUKINOV, E. PUGACHEVA, C. F. QI *et al.*, 2000 Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. Curr Biol **10:** 853-856.

KANDURI, C., G. FITZPATRICK, R. MUKHOPADHYAY, M. KANDURI, V. LOBANENKOV *et al.*, 2002 A differentially methylated imprinting control region within the Kcnq1 locus harbors a methylation-sensitive chromatin insulator. J Biol Chem **277:** 18106-18110.

KANEDA, M., M. OKANO, K. HATA, T. SADO, N. TSUJIMOTO *et al.*, 2004 Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature **429:** 900-903.

KANKEL, M. W., D. E. RAMSEY, T. L. STOKES, S. K. FLOWERS, J. R. HAAG *et al.*, 2003 Arabidopsis MET1 cytosine methyltransferase mutants. Genetics **163:** 1109-1122.

KELLY, W. G., and A. FIRE, 1998 Chromatin silencing and the maintenance of a functional germline in Caenorhabditis elegans. Development **125:** 2451-2456.

KHOSLA, S., A. AITCHISON, R. GREGORY, N. D. ALLEN and R. FEIL, 1999 Parental allele-specific chromatin configuration in a boundary-imprinting-control element upstream of the mouse H19 gene. Mol Cell Biol **19:** 2556-2566.

KHOSLA, S., G. MENDIRATTA and V. BRAHMACHARI, 2006 Genomic imprinting in the mealybugs. Cytogenet Genome Res **113:** 41-52.

KINOSHITA, T., R. YADEGARI, J. J. HARADA, R. B. GOLDBERG and R. L. FISCHER, 1999 Imprinting of the MEDEA Polycomb gene in the Arabidopsis endosperm. Plant Cell **11:** 1945-1952.

KINOSHITA, T., A. MIURA, Y. CHOI, Y. KINOSHITA, X. CAO *et al.*, 2004 One-way control of *FWA* imprinting in Arabidopsis endosperm by DNA methylation. Science **303:** 521-523.

KIYOSUE, T., N. OHAD, R. YADEGARI, M. HANNON, J. DINNENY *et al.*, 1999 Control of fertilization-independent endosperm development by the MEDEA polycomb gene in Arabidopsis. Proc Natl Acad Sci U S A **96:** 4186-4191.

KLEBES, A., A. SUSTAR, K. KECHRIS, H. LI, G. SCHUBIGER *et al.*, 2005 Regulation of cellular plasticity in Drosophila imaginal disc cells by the Polycomb group, trithorax group, and lama genes. Development **132:** 3753.

KLYMENKO, T., B. PAPP, W. FISCHLE, T. KOCHER, M. SCHELDER *et al.*, 2006 A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. Genes Dev **20:** 1110-1122.

KOBAYASHI, H., C. SUDA, T. ABE, Y. KOHARA, T. IKEMURA *et al*., 2006 Bisulfite sequencing and dinucleotide content analysis of 15 imprinted mouse differentially methylated regions (DMRs): paternally methylated DMRs contain less CpGs than maternally methylated DMRs. Cytogenet Genome Res **113:** 130-137.

KORF, I., Y. FAN and S. STROME, 1998 The Polycomb group in Caenorhabditis elegans and maternal control of germline development. Development **125:** 2469-2478.

KUNERT, N., J. MARHOLD, J. STANKE, D. STACH and F. LYKO, 2003 A Dnmt2-like protein mediates DNA methylation in Drosophila. Development **130:** 5083-5090.

KUZMICHEV, A., T. JENUWEIN, P. TEMPST and D. REINBERG, 2004 Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. Mol Cell **14:** 183-193.

KUZMICHEV, A., R. MARGUERON, A. VAQUERO, T. S. PREISSNER, M. SCHER *et al*., 2005 Composition and histone substrates of polycomb repressive group complexes change during cellular differentiation. Proc Natl Acad Sci U S A **102:** 1859-1864.

LABRADOR, M., and V. G. CORCES, 2002 Setting the boundaries of chromatin domains and nuclear organization. Cell **111:** 151-154.

LAGOS-QUINTANA, M., R. RAUHUT, W. LENDECKEL and T. TUSCHL, 2001 Identification of novel genes coding for small expressed RNAs. Science **294:** 853-858.

LAU, N. C., L. P. LIM, E. G. WEINSTEIN and D. P. BARTEL, 2001 An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science **294:** 858-862.

LAWRENCE, P. A., P. JOHNSTON and G. STRUHL, 1983 Different requirements for homeotic genes in the soma and germ line of Drosophila. Cell **35:** 27-34.

LEE, J., K. INOUE, R. ONO, N. OGONUKI, T. KOHDA *et al.*, 2002 Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. Development **129:** 1807-1817.

LEE, N., C. MAURANGE, L. RINGROSE and R. PARO, 2005 Suppression of Polycomb group proteins by JNK signalling induces transdetermination in Drosophila imaginal discs. Nature **438:** 234-237.

LEE, R. C., and V. AMBROS, 2001 An extensive class of small RNAs in Caenorhabditis elegans. Science **294:** 862-864.

LEWIS, A., K. MITSUYA, D. UMLAUF, P. SMITH, W. DEAN *et al.*, 2004 Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation. Nat Genet **36:** 1291-1295.

LEWIS, A., and W. REIK, 2006 How imprinting centres work. Cytogenet Genome Res **113:** 81-89.

LI, E., C. BEARD and R. JAENISCH, 1993 Role for DNA methylation in genomic imprinting. Nature **366:** 362-365.

LLOYD, V., 2000 Parental imprinting in Drosophila. Genetica **109:** 35-44.

LOBNER-OLESEN, A., O. SKOVGAARD and M. G. MARINUS, 2005 Dam methylation: coordinating cellular processes. Curr Opin Microbiol **8:** 154-160.

LOPES, S., A. LEWIS, P. HAJKOVA, W. DEAN, J. OSWALD *et al.*, 2003 Epigenetic modifications in an imprinting cluster are controlled by a hierarchy of DMRs suggesting long-range chromatin interactions. Hum Mol Genet **12:** 295-305.

LOPEZ-ALANON, D. M., L. A. LOPEZ-FERNANDEZ, V. CASTANEDA, D. B. KRIMER and J. DEL MAZO, 1997 H3.3A variant histone mRNA containing an alpha-globin

insertion: modulated expression during mouse gametogenesis correlates with meiotic onset. DNA Cell Biol **16:** 639-644.

LOPPIN, B., E. BONNEFOY, C. ANSELME, A. LAURENCON, T. L. KARR *et al.*, 2005 The histone H3.3 chaperone HIRA is essential for chromatin assembly in the male pronucleus. Nature **437:** 1386-1390.

LUCIFERO, D., M. R. MANN, M. S. BARTOLOMEI and J. M. TRASLER, 2004 Gene-specific timing and epigenetic memory in oocyte imprinting. Hum Mol Genet **13:** 839-849.

LYKO, F., B. H. RAMSAHOYE and R. JAENISCH, 2000 DNA methylation in Drosophila melanogaster. Nature **408:** 538-540.

LYLE, R., D. WATANABE, D. TE VRUCHTE, W. LERCHNER, O. W. SMRZKA *et al.*, 2000 The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1. Nat Genet **25:** 19-21.

LYNCH, C., and M. TRISTEM, 2003 A co-opted gypsy-type LTR-retrotransposon is conserved in the genomes of humans, sheep, mice, and rats. Curr Biol **13:** 1518-1523.

LYON, M. F., 1998 X-chromosome inactivation: a repeat hypothesis. Cytogenet Cell Genet **80:** 133-137.

MAGER, J., N. D. MONTGOMERY, F. P. DE VILLENA and T. MAGNUSON, 2003 Genome imprinting regulated by the mouse *Polycomb* group protein Eed. Nat Genet **33:** 502-507.

MANCINI-DINARDO, D., S. J. STEELE, R. S. INGRAM and S. M. TILGHMAN, 2003 A differentially methylated region within the gene Kcnq1 functions as an imprinted promoter and silencer. Hum Mol Genet **12:** 283-294.

MANCINI-DINARDO, D., S. J. STEELE, J. M. LEVORSE, R. S. INGRAM and S. M. TILGHMAN, 2006 Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. Genes Dev **20:** 1268-1282.

MARTIANOV, I., S. BRANCORSINI, R. CATENA, A. GANSMULLER, N. KOTAJA *et al.*, 2005 Polar nuclear localization of H1T2, a histone H1 variant, required for spermatid elongation and DNA condensation during spermiogenesis. Proc Natl Acad Sci U S A **102:** 2808-2813.

MARTIN, D. I., R. WARD and C. M. SUTER, 2005 Germline epimutation: a basis for epigenetic disease in humans. Ann N Y Acad Sci **1054:** 68-77.

MAYER, W., A. NIVELEAU, J. WALTER, R. FUNDELE and T. HAAF, 2000 Demethylation of the zygotic paternal genome. Nature **403:** 501-502.

MULLER, J., C. M. HART, N. J. FRANCIS, M. L. VARGAS, A. SENGUPTA *et al.*, 2002 Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell **111:** 197-208.

MURRELL, A., S. HEESON and W. REIK, 2004 Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. Nat Genet **36:** 889-893.

NEUMANN, B., P. KUBICKA and D. P. BARLOW, 1995 Characteristics of imprinted genes. Nat Genet **9:** 12-13.

NICKEL, B. E., S. Y. ROTH, R. G. COOK, C. D. ALLIS and J. R. DAVIE, 1987 Changes in the histone H2A variant H2A.Z and polyubiquitinated histone species in developing trout testis. Biochemistry **26:** 4417-4421.

O'CARROLL, D., S. ERHARDT, M. PAGANI, S. C. BARTON, M. A. SURANI *et al.*, 2001 The polycomb-group gene Ezh2 is required for early mouse development. Mol Cell Biol **21:** 4330-4336.

OHNO, M., N. AOKI and H. SASAKI, 2001 Allele-specific detection of nascent transcripts by fluorescence in situ hybridization reveals temporal and culture-induced changes in Igf2 imprinting during pre-implantation mouse development. Genes Cells **6:** 249-259.

OKADA, T., M. ENDO, M. B. SINGH and P. L. BHALLA, 2005 Analysis of the histone H3 gene family in Arabidopsis and identification of the male-gamete-specific variant AtMGH3. Plant J **44:** 557-568.

OKAMURA, K., Y. SAKAKI and T. ITO, 2005 Comparative genomics approach toward critical determinants for the imprinting of an evolutionarily conserved gene Impact. Biochem Biophys Res Commun **329:** 824-830.

ONO, R., H. SHIURA, H. ABURATANI, T. KOHDA, T. KANEKO-ISHINO *et al.*, 2003 Identification of a large novel imprinted gene cluster on mouse proximal chromosome 6. Genome Res **13:** 1696-1705.

ORLANDO, V., E. P. JANE, V. CHINWALLA, P. J. HARTE and R. PARO, 1998 Binding of trithorax and Polycomb proteins to the bithorax complex: dynamic changes during early Drosophila embryogenesis. Embo J **17:** 5141-5150.

OSWALD, J., S. ENGEMANN, N. LANE, W. MAYER, A. OLEK *et al.*, 2000 Active demethylation of the paternal genome in the mouse zygote. Curr Biol **10:** 475-478.

PARO, R., and B. ZINK, 1992 The Polycomb gene is differentially regulated during oogenesis and embryogenesis in *Drosophila melanogaster*. Mech. Dev. **40:** 37-46.

PARRISH, S., J. FLEENOR, S. XU, C. MELLO and A. FIRE, 2000 Functional anatomy of a dsRNA trigger: differential requirement for the two trigger strands in RNA interference. Mol Cell **6:** 1077-1087.

PARRISH, S., and A. FIRE, 2001 Distinct roles for RDE-1 and RDE-4 during RNA interference in Caenorhabditis elegans. Rna **7:** 1397-1402.

PAULSEN, M., and A. C. FERGUSON-SMITH, 2001 DNA methylation in genomic imprinting, development, and disease. J Pathol **195:** 97-110.

PETRONIS, A., A. D. PATERSON and J. L. KENNEDY, 1999 Schizophrenia: an epigenetic puzzle? Schizophr Bull **25:** 639-655.

PIRROTTA, V., 1997 PcG complexes and chromatin silencing. Curr Opin Genet Dev **7:** 249-258.

RACHMILEWITZ, J., R. GOSHEN, I. ARIEL, T. SCHNEIDER, N. DE GROOT *et al.*, 1992 Parental imprinting of the human H19 gene. FEBS Lett **309:** 25-28.

REIK, W., A. COLLICK, M. L. NORRIS, S. C. BARTON and M. A. SURANI, 1987 Genomic imprinting determines methylation of parental alleles in transgenic mice. Nature **328:** 248-251.

REIK, W., and J. WALTER, 2001 Genomic imprinting: parental influence on the genome. Nat Rev Genet **2:** 21-32.

REINHART, B., M. ELJANNE and J. R. CHAILLET, 2002 Shared role for differentially methylated domains of imprinted genes. Mol Cell Biol **22:** 2089-2098.

RIDGWAY, P., K. D. BROWN, D. RANGASAMY, U. SVENSSON and D. J. TREMETHICK, 2004 Unique residues on the H2A.Z containing nucleosome surface are important for Xenopus laevis development. J Biol Chem **279:** 43815-43820.

RILEY, P. D., S. B. CARROLL and M. P. SCOTT, 1987 The expression and regulation of Sex combs reduced protein in Drosophila embryos. Genes Dev **1:** 716-730.

RINGROSE, L., H. EHRET and R. PARO, 2004 Distinct contributions of histone H3 lysine 9 and 27 methylation to locus-specific stability of polycomb complexes. Mol Cell **16:** 641-653.

ROBERTSON, K. D., 2005 DNA methylation and human disease. Nat Rev Genet **6:** 597-610.

RONEMUS, M. J., M. GALBIATI, C. TICKNOR, J. CHEN and S. L. DELLAPORTA, 1996 Demethylation-induced developmental pleiotropy in Arabidopsis. Science **273:** 654-657.

ROSSANT, J., 1993 Immortal germ cells? Curr Biol **3:** 47-49.

ROYO, H., M. L. BORTOLIN, H. SEITZ and J. CAVAILLE, 2006 Small non-coding RNAs and genomic imprinting. Cytogenet Genome Res **113:** 99-108.

SANTOS, F., B. HENDRICH, W. REIK and W. DEAN, 2002 Dynamic reprogramming of DNA methylation in the early mouse embryo. Dev Biol **241:** 172-182.

SATO, S., T. YOSHIMIZU, E. SATO and Y. MATSUI, 2003 Erasure of methylation imprinting of Igf2r during mouse primordial germ-cell development. Mol Reprod Dev **65:** 41-50.

SCHWEIZER, J., D. ZYNGER and U. FRANCKE, 1999 In vivo nuclease hypersensitivity studies reveal multiple sites of parental origin-dependent differential chromatin conformation in the 150 kb SNRPN transcription unit. Hum Mol Genet **8:** 555-566.

SCOTT, R. J., and M. SPIELMAN, 2004 Epigenetics: imprinting in plants and mammals--the same but different? Curr Biol **14:** R201-203.

SCOTT, R. J., and M. SPIELMAN, 2006 Genomic imprinting in plants and mammals: how life history constrains convergence. Cytogenet Genome Res **113:** 53-67.

SHA, K., and A. FIRE, 2005 Imprinting capacity of gamete lineages in Caenorhabditis elegans. Genetics **170:** 1633-1652.

SHUMACHER, A., C. FAUST and T. MAGNUSON, 1996 Positional cloning of a global regulator of anterior-posterior patterning in mice. Nature **383:** 250-253.

SIJEN, T., J. FLEENOR, F. SIMMER, K. L. THIJSSEN, S. PARRISH *et al.*, 2001 On the role of RNA amplification in dsRNA-triggered gene silencing. Cell **107:** 465-476.

SIMPSON, V. J., T. E. JOHNSON and R. F. HAMMEN, 1986 Caenorhabditis elegans DNA does not contain 5-methylcytosine at any time during development or aging. Nucleic Acids Res **14:** 6711-6719.

SLEUTELS, F., R. ZWART and D. P. BARLOW, 2002 The non-coding *Air* RNA is required for silencing autosomal imprinted genes. Nature **415:** 810-813.

SPILLANE, C., C. BAROUX, J. M. ESCOBAR-RESTREPO, D. R. PAGE, S. LAOUEILLE *et al.*, 2004 Transposons and tandem repeats are not involved in the control of genomic imprinting at the MEDEA locus in Arabidopsis. Cold Spring Harb Symp Quant Biol **69:** 465-475.

SULSTON, J. E., E. SCHIERENBERG, J. G. WHITE and J. N. THOMSON, 1983 The embryonic cell lineage of the nematode *Caenorhabditis elegans*. Dev Biol **100:** 64-119.

SWAIN, J. L., T. A. STEWART and P. LEDER, 1987 Parental legacy determines methylation and expression of an autosomal transgene: a molecular mechanism for parental imprinting. Cell **50:** 719-727.

SZABO, P., S. H. TANG, A. RENTSENDORJ, G. P. PFEIFER and J. R. MANN, 2000 Maternal-specific footprints at putative CTCF sites in the H19 imprinting control region give evidence for insulator function. Curr Biol **10:** 607-610.

TAKIHARA, Y., D. TOMOTSUNE, M. SHIRAI, Y. KATOH-FUKUI, K. NISHII *et al.*, 1997 Targeted disruption of the mouse homologue of the Drosophila polyhomeotic gene leads to altered anteroposterior patterning and neural crest defects. Development **124:** 3673-3682.

TANIMOTO, K., M. SHIMOTSUMA, H. MATSUZAKI, A. OMORI, J. BUNGERT *et al.*, 2005 Genomic imprinting recapitulated in the human beta-globin locus. Proc Natl Acad Sci U S A **102:** 10250-10255.

THORVALDSEN, J. L., K. L. DURAN and M. S. BARTOLOMEI, 1998 Deletion of the *H19* differentially methylated domain results in loss of imprinted expression of *H19* and *Igf2*. Genes Dev **12:** 3693-3702.

TSAI, C. E., S. P. LIN, M. ITO, N. TAKAGI, S. TAKADA *et al.*, 2002 Genomic imprinting contributes to thyroid hormone metabolism in the mouse embryo. Curr Biol **12:** 1221-1226.

TUCKER, K. L., C. BEARD, J. DAUSMANN, L. JACKSON-GRUSBY, P. W. LAIRD *et al.*, 1996 Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes. Genes Dev **10:** 1008-1020.

TURKER, M. S., and T. H. BESTOR, 1997 Formation of methylation patterns in the mammalian genome. Mutat Res **386:** 119-130.

UMLAUF, D., Y. GOTO, R. CAO, F. CERQUEIRA, A. WAGSCHAL *et al.*, 2004 Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of polycomb group complexes. Nat Genet **36:** 1296.

VERONA, R. I., M. R. MANN and M. S. BARTOLOMEI, 2003 Genomic imprinting: intricacies of epigenetic regulation in clusters. Annu Rev Cell Dev Biol **19:** 237-259.

VIELLE-CALZADA, J. P., J. THOMAS, C. SPILLANE, A. COLUCCIO, M. A. HOEPPNER *et al.*, 1999 Maintenance of genomic imprinting at the Arabidopsis medea locus requires zygotic DDM1 activity. Genes Dev **13:** 2971-2982.

VIRE, E., C. BRENNER, R. DEPLUS, L. BLANCHON, M. FRAGA *et al.*, 2006 The polycomb group protein EZH2 directly controls DNA methylation. Nature **439:** 871-874.

VU, T. H., and A. R. HOFFMAN, 1994 Promoter-specific imprinting of the human insulin-like growth factor-II gene. Nature **371:** 714-717.

VU, T. H., T. LI and A. R. HOFFMAN, 2004 Promoter-restricted histone code, not the differentially methylated DNA regions or antisense transcripts, marks the imprinting status of IGF2R in human and mouse. Hum Mol Genet **13:** 2233-2245.

WALTER, J., B. HUTTER, T. KHARE and M. PAULSEN, 2006 Repetitive elements in imprinted genes. Cytogenet Genome Res **113:** 109-115.

WANG, H., L. WANG, H. ERDJUMENT-BROMAGE, M. VIDAL, P. TEMPST *et al.*, 2004 Role of histone H2A ubiquitination in Polycomb silencing. Nature **431:** 873-878.

WATSON, C. E., S. Y. GAUTHIER and P. L. DAVIES, 1999 Structure and expression of the highly repetitive histone H1-related sperm chromatin proteins from winter flounder. Eur J Biochem **262:** 258-267.

WUTZ, A., O. W. SMRZKA, N. SCHWEIFER, K. SCHELLANDER, E. F. WAGNER *et al.*, 1997 Imprinted expression of the *Igf2r* gene depends on an intronic CpG island. Nature **389:** 745-749.

XIAO, W., M. GEHRING, Y. CHOI, L. MARGOSSIAN, H. PU *et al.*, 2003 Imprinting of the MEA Polycomb gene is controlled by antagonism between MET1 methyltransferase and DME glycosylase. Dev Cell **5:** 891-901.

XIN, Z., C. D. ALLIS and J. WAGSTAFF, 2001 Parent-specific complementary patterns of histone H3 lysine 9 and H3 lysine 4 methylation at the Prader-Willi syndrome imprinting center. Am J Hum Genet **69:** 1389-1394.

YAMAZAKI, Y., E. W. LOW, Y. MARIKAWA, K. IWAHASHI, M. S. BARTOLOMEI *et al.*, 2005 Adult mice cloned from migrating primordial germ cells. Proc Natl Acad Sci U S A **102:** 11361-11366.

YOUNGSON, N. A., S. KOCIALKOWSKI, N. PEEL and A. C. FERGUSON-SMITH, 2005 A small family of sushi-class retrotransposon-derived genes in mammals and their relation to genomic imprinting. J Mol Evol **61:** 481-490.

YU, B. D., R. D. HANSON, J. L. HESS, S. E. HORNING and S. J. KORSMEYER, 1998 MLL, a mammalian trithorax-group gene, functions as a transcriptional maintenance factor in morphogenesis. Proc Natl Acad Sci U S A **95:** 10632-10636.

ZHANG, H., R. B. AZEVEDO, R. LINTS, C. DOYLE, Y. TENG *et al.*, 2003 Global regulation of Hox gene expression in *C. elegans* by a SAM domain protein. Dev Cell **4:** 903-915.

ZHANG, Y., R. CAO, L. WANG and R. S. JONES, 2004 Mechanism of Polycomb group gene silencing. Cold Spring Harb Symp Quant Biol **69:** 309-317.

CHAPTER 2

Part I: Forward genetic screen for *C. elegans* silencing mutants

INTRODUCTION

Every organism must face the prospect of invasion by nucleic acid-based parasites, such as viruses and transposable elements. Prokaryotes have evolved restriction-modification systems to defend against such parasites. Eukaryotes have evolved sophisticated gene silencing pathways that not only function in defense against selfish genetic elements, but also play roles in developmental programming of the organism. There is ample experimental evidence to support this assertion. In many organisms examined, knockout of factors involved in gene silencing processes leads to reactivation of previously dormant transposable elements in the genome or increased susceptibility to invasion by such parasites (HIROCHIKA *et al.* 2000; VASTENHOUW *et al.* 2003; WOODHOUSE *et al.* 2006). Introduction of foreign nucleic acids, either by a nucleic acid-based parasite or by a researcher, triggers pathways that lead to recognition and silencing of the foreign entity. The challenge for scientists has been to dissect the mechanism(s) of recognition and silencing.

Plasmids introduced into the *C. elegans* gonad have a propensity for concatamerization, forming long tandemly-repeated arrays (STINCHCOMB *et al.* 1985). In the first 1-2 generations, the transgene appears to undergo extensive rearrangement, but eventually stabilizes to a defined structure that varies little (if any) between subsequent generations (FIRE *et al.* 1991; MELLO *et al.* 1991; STINCHCOMB *et al.* 1985). Transgenes introduced by standard microinjection procedures tend to remain non-integrated (FIRE *et al.* 1991; MELLO *et al.* 1991; STINCHCOMB *et al.* 1985). Such extra-chromosomal transgenes are inherited faithfully through cell lineages but occasional loss of the

transgene does occur, resulting in mosaicism. In *C. elegans*, many extra-chromosomal transgenes above ≈700kb are transmitted at high frequencies (STINCHCOMB *et al.* 1985).

The structural complexity of a transgene array is a very strong determinant of its activity. Simple arrays, characterized by high repetitive character, usually exhibit poor expression, especially in the germline (FIRE and MELLO 1995). Improvements in transgene expression can be obtained by mixing sheared *C. elegans* genomic DNA with the desired transgene construct. The resulting "complex array" has lower repetitive character and lower copy number (FIRE and MELLO 1995). This relationship between copy number and expression level is clearly evident in *unc-54* rescue lines, where high-copy arrays tend to result in mosaic rescue (White-Harrison and Fire, unpublished).

Genetic screens for somatic silencing in other systems have led to the identification of factors that have dual roles in transgene silencing and development of the organism. Not surprisingly, the action of many of these factors results in chromatin remodeling, either directly or indirectly. Some examples include *ddm1* (JEDDELOH *et al.* 1999) and *mom* (AMEDEO *et al.* 2000) in *Arabidopsis*, *dim 5* (TAMARU and SELKER 2001) in *Neurospora*, and *Su*(*var*)2-5 (FANTI *et al.* 1998) in *Drosophila*. Many of these genes have homologs in *C. elegans*. However, transgene silencing has not been characterized in *C. elegans* to the extent that it has been investigated in other systems. There are certainly mutations in *C. elegans* that result in de-repression of transgene expression in somatic tissues. However, many of these genes were originally characterized in the context of other functions, and their transgene phenotypes remain largely unexplored. Examples of such mutations include *lin-61*(*az1*) (A. Zahler and R. Horvits, personal communication),

mut-7 (GRISHOK *et al*. 2000; KETTING *et al*. 1999), *pag-1* and *pag-2* (E. Aamodt, personal communication).

Hsieh *et al*. previously identified TAM-1 and LIN-35 as being involved in somatic silencing in *C. elegans* (HSIEH *et al*. 1999). TAM-1 and LIN-35 negatively regulate Ras pathway activities in *C. elegans*. Loss-of-function mutations in either gene result in hypersilencing of transgenes in a context-dependent manner: only highly-repetitive transgenes (simple arrays) are affected in a *tam-1* loss-of-function background. In addition to having a transgene phenotype, *tam-1* loss-of-function also has mild developmental phenotypes.

We sought to extend this work by carrying out additional genetic screens for factors that modulate transgene activities. In particular, we were interested in endogenous *C. elegans* factors whose loss-of-function mutations lead to de-silencing of transgenes. Such mutants would be useful for transgenic rescue or analysis in which high activity of the transgene is desired. They may also identify factors involved in regulatory chromatin structure or gene expression. We constructed a transgenic line (PD3815) carrying a GFP reporter driven by the *unc-54* muscle promoter. The extra-chromosomal simple array in PD3815, designated *ccEx3815*, is silenced. PD3815 animals show weak and mosaic expression of the GFP reporter. Multiple EMS mutageneses led to the isolation of a handful of candidates including three (PD3852, PD3861, and PD3862) that we characterized in detail. These three candidates lacked second-site mutations that we searched for, but nevertheless displayed some interesting properties, including integration and de-silencing of the transgene, presence of a parent-of-origin effect in two of the three

candidates, and non-linear GFP expression in array homozygotes compared to array hemizygotes.

MATERIALS AND METHODS

C. elegans strains and growth conditions

Worms were reared on *E. coli* strain OP50 grown on NGM (nematode growth medium) nutrient plates according to standard protocols (BRENNER 1974). Animals were reared at 23°C unless stated otherwise. All genetic crosses were performed at 23°C unless specifically stated otherwise. Worm strains used in the experiments were as follows:

N2: wildtype strain of *C. elegans* (Bristol isolate)

***pha-1*(*e2123ts*) III:** carries the temperature-sensitive recessive mutation *pha-1*(*e2123ts*) on Chromosome III (SCHNABEL and SCHNABEL 1990); homozygous *pha-1*(*e2123ts*) animals are viable at 16°C but embryonic lethal at temperatures above 20°C

***edIs6*[*unc-119::gfp*]:** *unc-119*::gfp translational fusion (MADURO and PILGRIM 1995)

PD3815 [*pha-1*(*e2123ts*) III; *ccEx3815*]: carries the extra-chromosomal tandem array *ccEx3815*[*unc-54::gfp* + *pha-1*(+)] in the *pha-1*(*e2123ts*) background; there are actually three different *unc-54::gfp* constructs, one localized to the nucleus, one to the mitochondria, and one to the nucleolus

PD3816 [*pha-1*(*e2123ts*) III; *him-5*(*e1467*) V]: carries the *pha-1*(*e2123ts*) mutation (see above) as well as the *him-5*(*e1467*) mutation which induces a high frequency of male progeny

PD3819[*pha-1*(*e2123ts*) III; *ccIn3861* V]: PD3861 outcrossed four times

PD3852 [*ccIn3852 pha-1*(*e2123ts*) III]: an integrated derivative of *ccEx3815*. The *unc-54::gfp* transgene is integrated into chromosome III; this strain is in the *pha-1*(*e2123ts*) background

PD3854 [*ccIn3852 pha-1*(*e2123ts*) III]: This strain is PD3852 outcrossed to N2 once

PD3861 [*pha-1*(*e2123ts*) III; *ccIn3861* V]: an integrated derivative of *ccEx3815*. The *unc-54::gfp* transgene is integrated into chromosome V; this strain is in the *pha-1*(*e2123ts*) background

PD3862 [*ccIn3862* I; *pha-1*(*e2123ts*) III]: an integrated derivative of *ccEx3815*. The *unc-54::gfp* transgene is integrated into chromosome I; this strain is in the *pha-1*(*e2123ts*) background

PD3870 [*ccIn3870* ?; *pha-1*(*e2123ts*) III]: an integrated derivative of *ccEx3815*. The *unc-54::gfp* transgene integration site has not been determined; this strain is in the *pha-1*(*e2123ts*) background

PD3872 [*ccIn3852 pha-1*(*e2123ts*) III]: PD3852 outcrossed once

PD3873 [*ccIn3852 dpy-17*(*e164*) *unc-32*(*e189*) *pha-1*(*e2123ts*) III; *him-5*(*e1467*) V]: *ccIn3852* marked *in cis* with the two recessive markers *dpy-17*(*e164*) and *unc-32*(*e189*) and *in trans* with the recessive marker *him-5*(*e1467*)

PD3891 [*ccIn3891*; *pha-1*(*e2123ts*) III]: an integrated derivative of *ccEx3815*. The *unc-54::gfp* transgene integration site has not been determined; this strain is in the *pha-1*(*e2123ts*) background

PD3924 [*ccIn3862 dpy-5*(*e61*) *unc-13*(*e1091*) I; *pha-1*(*e2123ts*) III]: PD3862 marked *in cis* with the two recessive mutations *dpy-5*(*e61*) and *unc-13*(*e1091*)

PD7280 [*pha-1*(*e2123ts*) *dpy-17*(*e164*) *unc-32*(*e189*) III]: *pha-1*(*e2123ts*) marked *in cis* with the two recessive markers *dpy-17*(*e164*) and *unc-32*(*e189*)

PD7281 [*dpy-5*(*e61*) *unc-54*(*e1091*) I ; *pha-1*(*e2123ts*) III]: *pha-1*(*e2123ts*) marked *in trans* with the two recessive markers *dpy-5*(*e61*) and *unc-54*(*e1091*); used for mapping

PD7282 [*dpy-10*(*e128*) *unc-4*(*e120*) II; *pha-1*(*e2123ts*) III]: *pha-1*(*e2123ts*) marked *in trans* with the two recessive markers *dpy-10*(*e128*) *unc-4*(*e120*); used for mapping

PD7284 [*dpy-4*(*e1166*) *unc-17*(*e245*) IV; *pha-1*(*e2123ts*) III]: *pha-1*(*e2123ts*) marked *in trans* with the two recessive markers *dpy-4*(*e1166*) *unc-17*(*e245*); used for mapping

PD7285 [*dpy-11*(*e224*) *unc-60*(*e723*) V; *pha-1*(*e2123ts*) III]: *pha-1*(*e2123ts*) marked *in trans* with the two recessive markers *dpy-11*(*e224*) *unc-60*(*e723*); used for mapping

Plasmids used to establish transgenic lines

pC1: contains the wildtype genomic *pha-1* sequence without the 3' UTR (GRANATO *et al*. 1994). The genomic *pha-1* sequences is used as a transformation marker.

pPD95.93: carries a 204bp *unc-54* promoter segment driving the GFP coding region followed by the *unc-54* 3' UTR. pPD95.93 also carries a nuclear localization signal and *lacZ* (yielding nuclear GFP)

pPD105.21: carries a 204bp *unc-54* promoter segment driving the GFP coding region followed by the *unc-54* 3' UTR. pPD105.21 also carries a mitochondrial localization signal (yielding mitochondrial GFP)

pPD120.90: carries a 204bp *unc-54* promoter segment driving the GFP coding region followed by the *unc-54* 3' UTR. pPD120.90 also carries four nuclear localization signals (yielding nucleolar GFP)

Construction of transgenic animals

Two mixtures of the four plasmids pC1, pPD95.93, pPD105.21, and pPD129.90 were micro-injected into *pha-1*(*e2123ts*) worms according to standard procedures (MELLO *et al*. 1991). Each injection mix differed only in the concentration of plasmids. Injection mix #1 contained 20ng each of pPD95.93, pPD105.21, pPD129.90, and 750ng of pC1 in a total volume of 6μL. This mix gave rise to line JF3067. Injection mix #2 contained 133ng each of pPD95.93, pPD105.21, pPD129.90, and 750ng of pC1 in a total volume of 6μL. This injection mix gave rise to lines JF3070 and JF3071. Following micro-injection, animals were reared at 23°C to select for transformants. Only animals that harbor the transgene are viable at 23°C.

Microscopy

Observations of worms were made using dissecting and compound microscopes fitted with GFP filters. To take digitized pictures, animals were immobilized in mounting solution (50 mM NaCl, 5 mM EDTA, 0.5 mM levamisole) and images digitized using a CCD camera (Nikon CCD300ET-RC camera) mounted on the microscope.

EMS mutagenesis

Mutagenesis was carried out using standard protocols. Non-starved PD3815 animals were washed off feeding plates with M9 Buffer (22mM KH_2PO_4, 42mM Na_2HPO_4, 86mM NaCl, 1mM $MgSO_4$), spun down to concentrate animals, and animals

transferred to EMS solution in the concentration range 25-50µg/mL. After mutagenesis, animals were washed five times with M9 buffer and transferred to fresh feeding plates.

RESULTS

To screen for endogenous factors that silence transgene expression in *C. elegans*, we engineered a transgenic line harboring a silenced GFP reporter. This line, PD3815, carries an extra-chromosomal transgene GFP reporter driven by the *unc-54* promoter. The *unc-54* gene, coding for a myosin heavy chain, is a component of the *C. elegans* bodywall musculature. For unknown reasons, *unc-54* simple transgene arrays are susceptible to silencing in transgenic animals (Fire, unpublished).

Construction of a transgenic line carrying a silenced GFP reporter and pedigree analysis

Three independent transformed lines were initially established by microinjection of plasmid mixes into *pha-1*(*e2123ts*) animals. The two plasmid mixes had the same plasmid compositions (pC1, pPD95.93, pPD105.21, and pPD129.90) but differed in the concentration of the plasmids (see Materials and Methods). Three progenitor lines were obtained and designated JF3067, JF3070, and JF3071. We traced the lineage of each progenitor line over at least five generations, keeping track of three variables: (1) penetrance: what percent of the population showed GFP expression?, (2) expressivity: what is the degree of GFP variation among individuals in the population?, and (3) mosaicism: how mosaic are individual animals in each lineage?

To obtain a transgenic line on which to use in the mutant screen, we performed a pedigree analysis on the three progenitor populations JF3067, JF3070, and JF3071. Figure 2.1 depicts how the analysis was performed for JF3067 (analyses for JF3070 and JF3071 were performed in exactly the same procedure). In particular, we were looking for a line that had the following characteristics: homogeneity in the population but some

degree of mosaicism in individual animals and low expression of the *unc-54::gfp*. The pedigree analysis essentially consisted of establishing subpopulations from single progenitors. At the JF3067.7.2.x level, we tracked the GFP expression of each subpopulation over five generations (Figure 2.2). Each F1 population was founded by a single JF3067.7.2.x animal (as such, the F1 generation always contained the fewest number of progeny). From the F1 generation, we observed the GFP profile of 20 randomly selected animals of the same stage and allowed these animals to be the founder of the F2 generation. We then randomly picked 20 F2 animals for observation and allowed them to be founders of the F3 generation, and so on for five generations. Several observations can be made from such analyses. First, the three progenitor populations JF3067, JF3070, and JF3071, which differed only in the concentration of plasmids they received, gave rise to distinctly different lineages with regards to homogeneity and intensity of GFP expression. For example, the JF3067 lineage shows mostly dim to GFP-negative animals (GFP-negative animals still harbored the transgene array, as these animals were *pha-1*(*e2123ts*) selected), while the JF3070 lineage showed even weaker GFP expression (Figure 2.2B). The JF3071 lineage was consistently less homogeneous compared to JF3067 and JF3070. In all the JF3071 subpopulations (Figure 2.2C), there was a large range of expression, from very bright expressers to very weak expressers. This difference in each of the lines was probably due to differences in initial plasmid concentration used in microinjections to establish each line.

Another observation is that the character of the population (range of expression and homogeneity) can only be partially transmitted to the next immediate generation. Such a character of the population is stochastic, not genetic. For example, less

homogeneous F1 populations (i.e. JF3067.7.2.7 Figure 2.2A) tend to give rise to less homogeneous subpopulations; whereas relatively homogeneous F1 populations (i.e. JF3067.2.8 Figure 2.2A) tend to produce lineages with more uniform and consistent GFP expression. The reason for this is clear when we look at populations derived from single founders. In Figure 2.3, each population was derived from a single JF3067.7.2.8, JF3067.7.2.9, or JF3067.7.2.10 founder. Note that, in general, GFP-positive founders tended to give rise to more GFP-positive progeny; while GFP-negative founders tended to produce more GFP-negative progeny.

Isolation of candidate mutants with the *bright* phenotype

After extensive pedigree analysis, we chose population JF3067.7.2.10 for use in the mutant screen. This population, renamed PD3815, carries a simple extra-chromosomal array designated *ccEx3815*. PD3815 animals exhibit weak GFP expression, but have some degree of homogeneity in the population (Figure 2.4, bottom panel). Additionally, there is a relatively high degree of mosaic GFP expression in individual animals (Figure 2.4, top panel). We expect that a mutant that fails to silence *ccEx3815* would exhibit strong, uniform GFP expression throughout the entire bodywall musculature of the animal (Figure 2.5). We call this phenotype "*bright*".

We performed mutageneses using standard EMS mutagenesis protocols and carried out both clonal and non-clonal screens according to the strategy depicted in Figure 2.6. Several mutagenized L4 (F0) animals were transferred to a plate and allowed to produce F1 progeny. For the clonal screen, we allowed individual F1 animals to establish a population of F2-F5 animals, a stage that should enrich for a recessive allele if one existed in the population. For the non-clonal screen, we scanned F0 plates for

possible mutant candidates. We performed a total of six mutagenesis and screened a total of 6,820 haploid genomes in the clonal screen. Five candidates show high penetrance and strong, uniform GFP expression (Figure 2.7). Seven additional candidates show weaker phenotypes (not as highly penetrant and/or weaker GFP expression). We also isolated six candidates that appear to be embryonic lethal. In this class of candidates, the embryos exhibit strong GFP expression but many do not hatch or hatch but die as L1/L2 larvae; the adults do not show the *bright* phenotype. Table 2.1 summarizes the five strongest *bright* candidates. We chose three candidates (PD3852, PD3861, and PD3862) for further analysis. PD3861 and PD3862 do not exhibit any perceptible developmental or morphological phenotype. PD3852 exhibits a "lagging *unc*" phenotype: in some outcrossed populations of PD3852, the population becomes progressively *unc* as it is passaged over multiple generations. Severe *unc*ness results if the population is not outcrossed again. The phenotype may or may not be present in all outcrossed populations, but when present, it is highly penetrant.

Integration of *ccEx3815* in at least five mutant candidates

Initial three-factor crosses to map the mutant locus in PD3852 led to the realization that the transgene array in PD3852, designated *ccIn3852*, had integrated into the middle of Chromosome III. Figures 2.8A and B show normal mapping situations of a recessive mutation (designated *m3852*) which de-silences an extra-chromosomal transgene array. Figure 2.8C shows the actual situation resulting from the mapping experiment, which deviated from both hypothetical mapping situations. Briefly, the candidate was crossed (as males) to a *pha-1*(*e2123ts*) mapping line marked with *dpy-17 unc-32* (designated *du* in Figure 2.8C). Upon selfing, all the F1 parents produced

only two classes of progeny: *bright* non-*du* and wildtype. Selfing 55 wildtype F2 animals resulted in 50 plates that had the same progeny classes as the F1 selfing and 5 recombinant plates containing *dpy* worms. When 33 *dpy* non-*bright* F3 worms were selfed, all produced the same two progeny classes seen in the F2 generation. The inset shows an interpretation of the genetic mapping data. There must have been an integration of the transgene array into Chromosome III near the *du* region. The initial F0 crossed would produce *pha-1 du* /*A* F1 animals; so that selfing F1 animals resulted in the two F2 classes with genotypes *pha-1 du* /*A* (wildtype) and *A*/*A* (*bright* non-*du*). Selfing of wildtype F2 animals resulted in mainly parental genotypes, but five plates contained recombinant progeny. Selfing of F3 *dpy* non-*bright* worms could only have given rise to *dpy bright* plus *dpy* non-*bright* animals in the F4 generation. These results also indicated that *ccIn3852* had integrated closer to *unc-32* than to *dpy-17*. Similar analyses with PD3861 and PD3862 revealed that the transgene arrays had been integrated in these lines as well.

We unequivocally showed an integration event in PD3852 by demonstrating that 100% of PD3852 animals failed to lose the transgene array when they were reared at 16°C. Since the transgenic animals are *pha-1*(*e2123ts*)-rescued by the transgene array at 23°C, animals which do not harbor the array at 23°C (the non-permissive temperature for *pha-1(e2123ts)*) hatch but die at the L1 stage. At 16°C (the permissive temperature for *pha-1(e2123ts)*), *pha-1*(*e2123ts*) animals do not require a rescuing transgene. Hence, the absence of array-negative PD3852 animals in 16°C (indicated by complete absence of GFP), showed that the transgene array in PD3852 was integrated. Similar analyses showed that the transgene array had been integrated in PD3861, PD3862, PD3870, and

PD3891. These integrated arrays are designated *ccIn3861*, *ccIn3862*, *ccIn3870*, and *ccIn3891*, respectively. Array *ccIn3852*, *ccIn3861*, and *ccIn3862* map to the centers of Chromosome III, Chromosome V, and Chromosome 1, respectively. The latter three lines were chosen for further analyses.

ccEx3815, *ccIn3852*, and *ccIn3862* exhibit a parent-of-origin effect

In outcrossing experiments, it was noticed that reciprocal crosses were not equivalent. In particular, certain integrated alleles of *ccEx3815* exhibited a parent-of-origin effect. Outcrosses in which the transgene was inherited from the male germline produced, on average, brighter F1 progeny than if the transgene was inherited from the oocyte. This was true for *ccIn3852*, *ccIn3862*, and *ccEx3815* itself but not for *ccIn3861*. *ccIn3870* and *ccIn3891* were not checked for a parent-of-origin effect. In addition to an imprinting effect, *ccIn3852*, *ccIn3861*, and *ccIn3862* show non-linear expression: animals homozygous for the transgene array show, on average, greater than two-fold GFP expression compared to animals hemizygous for the same array. The parent-of-origin effect and non-linear expression are discussed in detail in Chapter 3.

The *bright* phenotype in PD3852, PD3861, and PD3862 animals is due to integration of the transgene

To determine whether the *bright* phenotype in PD3852, PD3861, and PD3862 was due to a mutant locus, we performed the genetic experiment shown in Figure 2.9. In this experiment, we assumed that the hypothetical mutant locus, designated *m*, was not linked to the array. Nine F2 genotype classes result from the F0 cross and F1 selfing (Figure 2.9A). Those that do not harbor the array are lethal. Those that are homozygous for the transgene array are *bright* regardless of the presence of *m*. And those that are

hemizygous for the array are *bright* only if they are also homozygous for *m*. If *A*/+;*m*/*m* animals are allowed to self at 16°C, half their progeny will be GFP-negative; but all will be homozygous for *m*. If *m* does exist, then introduction of *ccEx3815* into GFP-negative animals would result in a population of *bright* worms.

The task was to distinguish between *A*/+;*m*/*m* animals from array homozygotes. *bright* F2 animals selfed at 16°C resulted in populations that were 100% GFP-positive (*A*/*A* founder) or populations of mixed GFP-positive and GFP-negative animals (*A*/+ founder). Introduction of *ccEx3815* into GFP-negative (*pha-1*/*pha-1*; *m*/*m*) animals did not result in any significant difference compared to the control (*ccEx3815* introduced into *pha*/*pha*; +/+ animals). The result of this experiment pointed to the lack of a second-site mutation in PD3852, PD3861, and PD3862.

A second line of evidence for the lack of second-site mutations in PD3852, PD3861, and PD3862 came from outcrossing experiments. In constructing genetically-marked derivatives of these three lines, at no time did we observe any non-*bright* animals. If the *bright* phenotype were due to a recessive second-site mutation, then the locus would have segregated away from the array during multiple outcrossings, and we would have observed animals that were GFP-positive, but not *bright* (i.e. like PD3815).

ccEx3815 and its integrated derivatives are structurally identical

To determine the structural complexities of *ccEx3815* and its integrated derivates, we performed a Southern hybridization, using as a probe a segment from the *unc-54* promoter. There does not appear to be a change in the structure of the transgene between the extra-chromosomal *ccEx3815* and the integrated *ccIn3852*, *ccIn3861*, and *ccIn3862* (Figure 2.10). However, it is clear that the *unc-54* promoter copy-number is far greater in

the transgenes than in wildtype (Figure 2.10B, arrow). Our estimates put the copy number of the *unc-54* promoter in *ccEx3815* and its derivatives to be about 20-30 copies.

DISCUSSION

We have attempted a screen for somatic silencing in *C. elegans*. We constructed a transgenic line (PD3815) carrying a mixture of *unc-54::gfp* plasmids with localization signals to the nucleus, mitochondria, and nucleolus. The extra-chromosomal transgene array in PD3815, designated *ccEx3815*, is silenced and exhibits mosaic expression (Figure 2.4). From a relatively small number of haploid genomes screened for a silencing phenotype, we isolated a handful of candidates, three of which we characterized in detail (PD3852, PD3861, and PD3862). Upon detailed analyses, we discovered that these three candidates (and possibly others as well) lack second-site mutations that conferred the *bright* phenotype. These false positives, however, exhibit some unexpected properties. First, integrated derivatives of *ccEx3815*, designated *ccIn3852*, *ccIn3861*, and *ccIn3862*, are apparently de-silenced; however, the structures of the transgene arrays remain identical in all four lines. Second, *ccIn3852*, *ccIn3862*, and *ccEx3815* (but not *ccIn3861*) exhibit a parent-of-origin effect. Third, *ccIn3852*, *ccIn3861*, and *ccIn3862* exhibit non-linear GFP expression between array hemizygotes and array homozygotes. The results of our screen thus indicate that the state of a transgene's activity is apparently context-dependent. Extensive studies in *C. elegans* and other model systems have shown that certain contexts (i.e. heterochromatin, position effects, highly repetitive character, copy number) inhibit expression while others (i.e. euchromatin, low-copy number, etc.) facilitate expression. The context can be determined by endogenous factors that act upon the transgene (i.e. chromatin remodeling factors that either repress or facilitate expression of the transgene) or by virtue of chance (i.e. integration site).

Integration of extra-chromosomal transgenes by EMS mutagenesis is rare in *C. elegans* (FIRE *et al.* 1991; MELLO *et al.* 1991), though not unprecedented. The preferred method to induce transgene integration in *C. elegans* is irradiation, which causes chromosome fragmentation and rearrangement, resulting in deletions, inversions, translocations, etc. (ANDERSON 1995). The DNA repair activities that follow such events likely incorporate the transgene into chromosomes. Thus, we were surprised to have recovered four (and probably more) integration events in a relatively small scale EMS screen. EMS (ethylmethanesulfonate) is an alkylating agent, acting by transferring an ethyl group to guanine. Ethylated guanine has a high propensity for mispairing with thymine. Upon DNA replication, the GT mismatch leads to a GC→AT transition (GRIFFITHS *et al.* 2000). Though primarily used to generate point and nonsense mutations, EMS has also been used to generate deletions (JANSEN *et al.* 1997). Presumably, any transgene array large enough to be recognized as chromosomal fragments would be incorporated into chromosomes by the double-stranded break repair pathway.

Transgene activity apparently is context-dependent in many systems examined thus far. There are a number of parameters which could affect the activity of a transgene. These include copy number and repetitive character, pairing state, position effects, DNA methylation state (HSIEH and FIRE 2000), histone modifications (COSGROVE and WOLBERGER 2005; MARTIN and ZHANG 2005; WOOD *et al.* 2005), insulator activities, and nuclear organization (CAPELSON and CORCES 2004; LABRADOR and CORCES 2002). A combination of these parameters (with the exception of DNA methylation, which is absent in *C. elegans*) is likely to be responsible for modulating the activity of *ccEx3815*

and its integrated derivatives. We note that not all integration events lead to a change in activity of the transgene, however. A similar *unc-54::gfp* transgene (*ccIn9385*) integrated into the genome remained inactive (HSIEH *et al*. 1999).

We note that all three integrated derivatives of *ccEx3815* are located at the centers of three distinct autosomes. We do not know the significance (if any) of transgene integration at these sites. There is not enough collective data in the *C. elegans* community to compare transgene activity between disparate integration sites. Certainly, there are features of *C. elegans* chromosomes that may modulate the activity of genes in that region. In general, the autosomal arms contain more repetitive character, have a higher recombination rate, and are relatively gene-poor compared to the central regions of the autosomes (BARNES *et al*. 1995; BRENNER 1974; CONSORTIUM 1998; GREENWALD *et al*. 1987; PRASAD and BAILLIE 1989; STARR *et al*. 1989). Whatever changes (if any) the extra-chromosomal transgene *ccEx3815* might have undergone, they led to a change in expression state in the integrated derivatives. It is interesting that *ccEx3815* did not undergo any rearrangements upon integration into the genome, as indicated by Southern blots (Figure 2.10). Perhaps EMS is a milder inducer of DNA double-stranded breaks, allowing *ccEx3815* to integrate into the genome while remaining intact. Thus, integration site and/or changes in copy number (our Southern blot is not quantitative and cannot distinguish any changes in copy number) may have played a role in modulating the activity of our transgene arrays.

Although screens for somatic silencing have been extensively reported in other model organisms, only a few such screens in *C. elegans* have been reported, possibly due to the intractable nature of such screens that we have encountered ourselves. In our

screen identify factors that modulate the activity of a silenced GFP reporter, *ccEx3815*, we unexpectedly selected for integration events of the reporter transgene. Surprisingly, the integrated derivatives of *ccEx3815* (*ccIn3852*, *ccIn3861*, *ccIn3862*) are structurally identical to the extra-chromosomal *ccEx3815*, but show improved expression over *ccEx3815*. An intriguing property is that the activity of *ccEx3815*, *ccIn3852*, and *ccIn3862* (but not *ccIn3861*) can be modulated in a parent-of-origin manner.

FIGURES

Figure 2.1. Pedigree analysis of JF3067, JF3070, and JF3071. The diagram shows an example analysis for JF3067, but JF3070 and JF3071 were analyzed in the same manner.

Figure 2.1

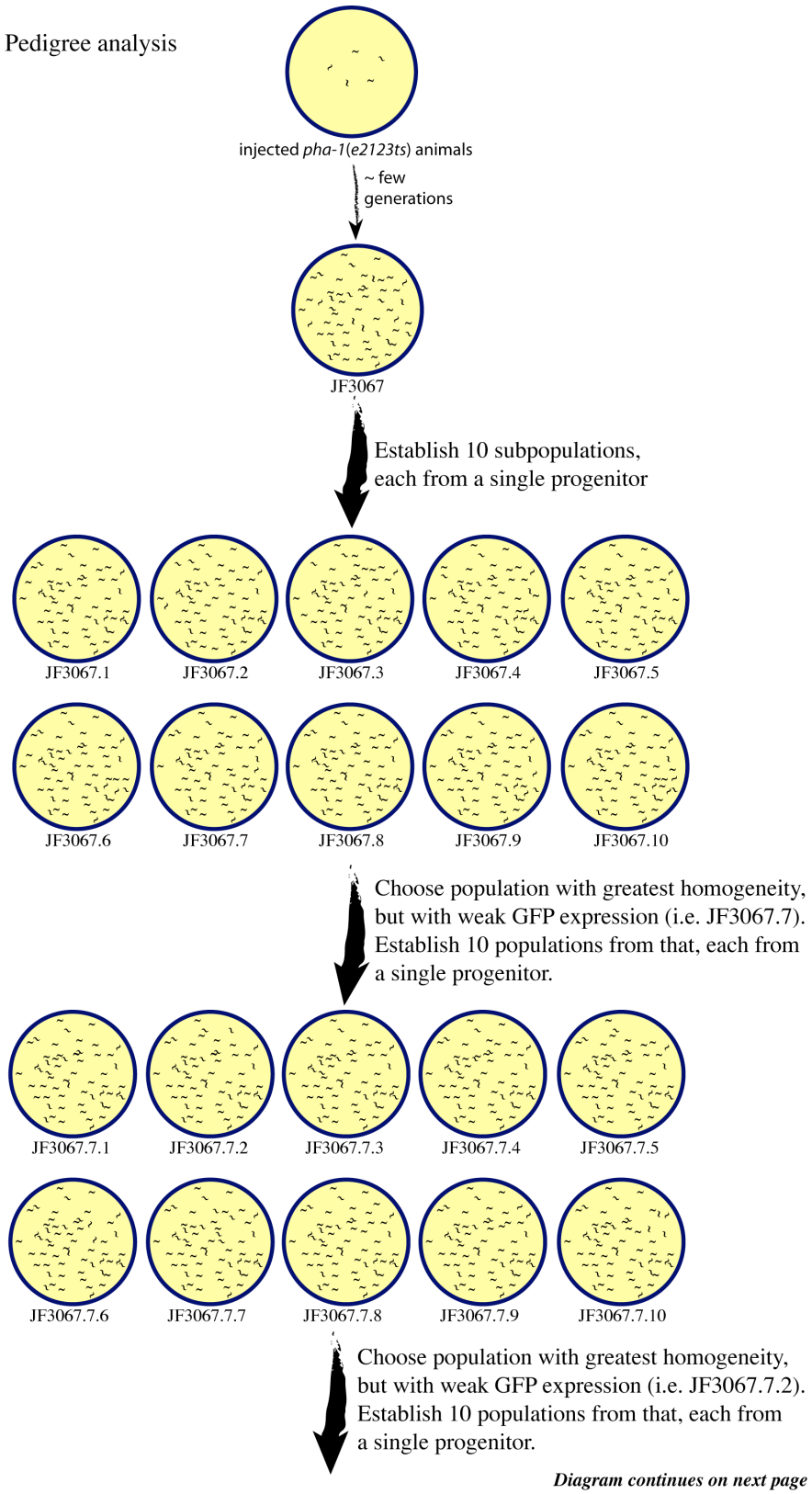
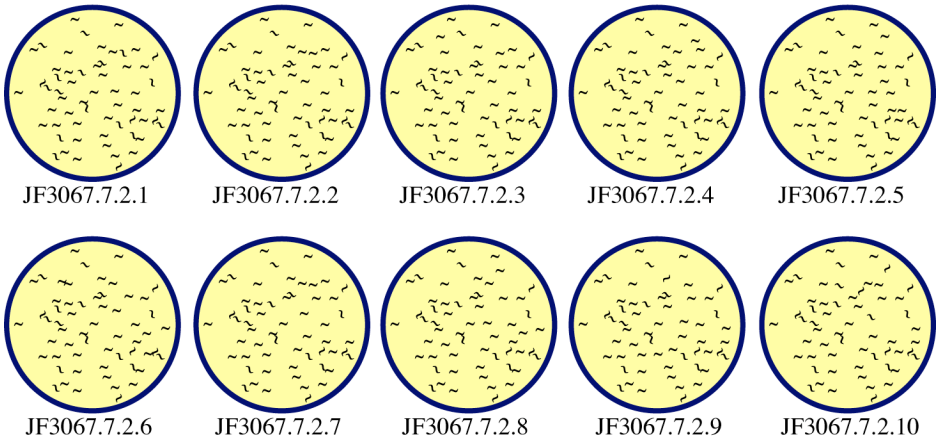


Diagram continues on next page

Figure 2.1 (*continued*)

pedigree analysis (*continued*)



For each JF3067.7.2.x population, pick one animal and allow it to self. Observe GFP intensity of the (F1) progeny. Let the F1 animals be the founder of the F2 generation.

For each JF3067.7.2.x population, randomly pick 20 F2 animals of the same stage (i.e. L4) and observe GFP intensity under the dissecting microscope. Allow the 20 F2 animals to be founders of next (F3) generation. Repeat until reach F5 for each JF3067.7.2.x population.

Based on pedigree analysis, choose a line (i.e. JF3067.7.2.10), with highest homogeneity in population, but with weak and mosaic GFP expression in individuals. Use this line for silencing screen.

Figure 2.2. Analysis of GFP variation among subpopulations derived from a single progenitor. From each of JF3067.7.2.x (part A), JF3070.4.2.x (part B), and JF3071.7.1.x (part C) lineage, we selfed a single (F0) progenitor. Approximately 10 F1 animals were randomly picked from the F1 generation, without knowledge of their GFP expression. We then scored the GFP expression of the randomly chosen animals under a dissecting microscope fitted with GFP filters. These F1 animals were allowed to found the F2 generation. Approximately 20 F2 animals were randomly and blindly picked from the population, and their GFP expression scored. This process was reiterated until the F5 generation. The degree of *bright*ness (dim, medium, *bright*) was based on subjective determination after weeks of familiarity with each of the lineages.

Figure 2.2

A

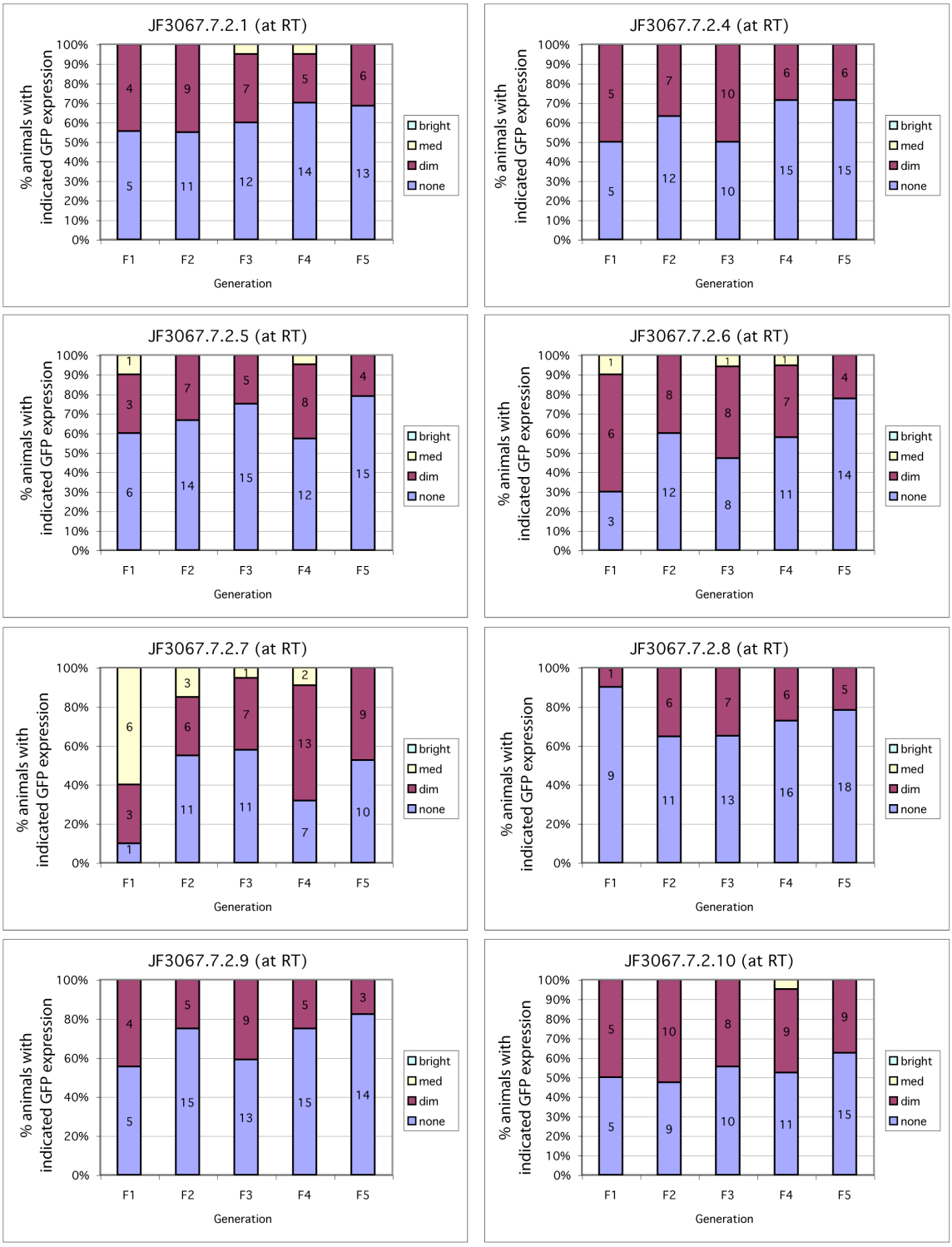

JF3067.7.2.1 (at RT)
JF3067.7.2.4 (at RT)
JF3067.7.2.5 (at RT)
JF3067.7.2.6 (at RT)
JF3067.7.2.7 (at RT)
JF3067.7.2.8 (at RT)
JF3067.7.2.9 (at RT)
JF3067.7.2.10 (at RT)
% animals with indicated GFP expression
Generation
F1
F2
F3
F4
F5
bright
med
dim
none

Figure 2.2 (*continued*)

B

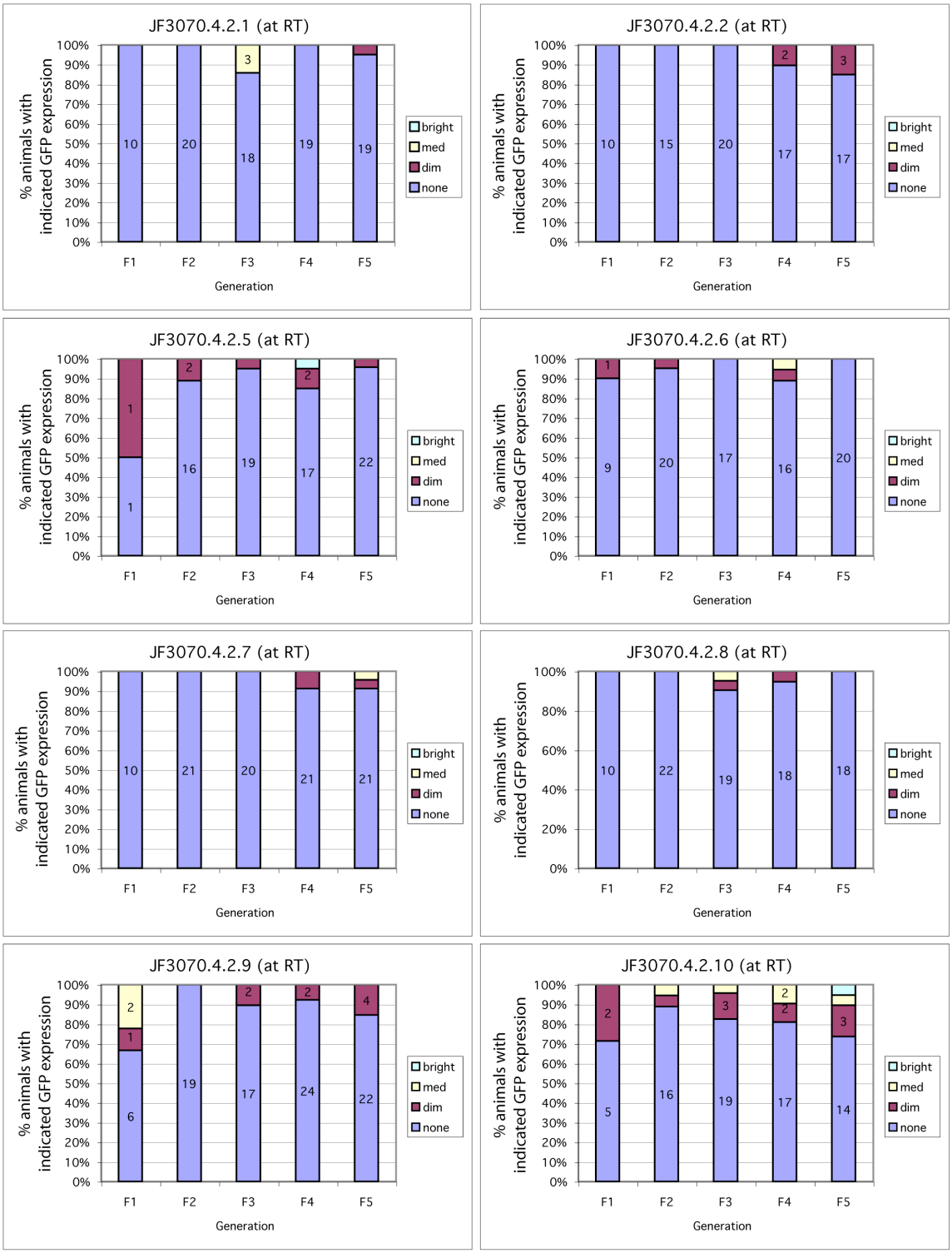


Figure 2.2 (*continued*)

C

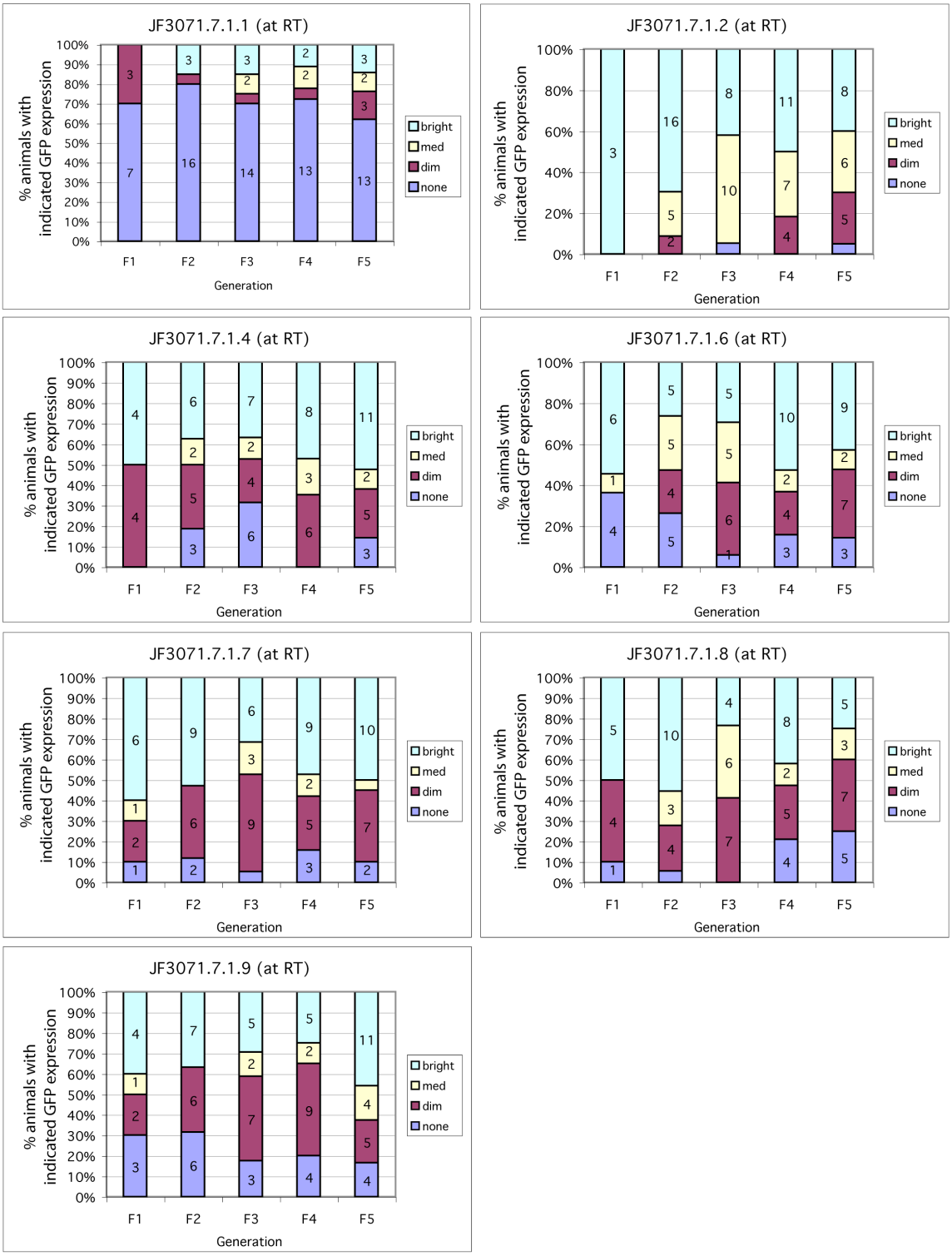


Figure 2.3. Stochasticity of GFP expression in PD3815 populations. Approximately ten animals from each JF3067.7.2.y population were randomly picked. These animals, labeled x=1, x=2, etc., were single founders of a population in which we observed the GFP profile. For example, from population JF3067.7.2.10, we selfed 10 animals (x=1, x=2, etc.) and then scored the GFP expression of each of the 10 populations. We always chose two groups of parents: weakly expressing parents (left of dashed line) and GFP-negative (but still array-positive) parents (right of dashed line).

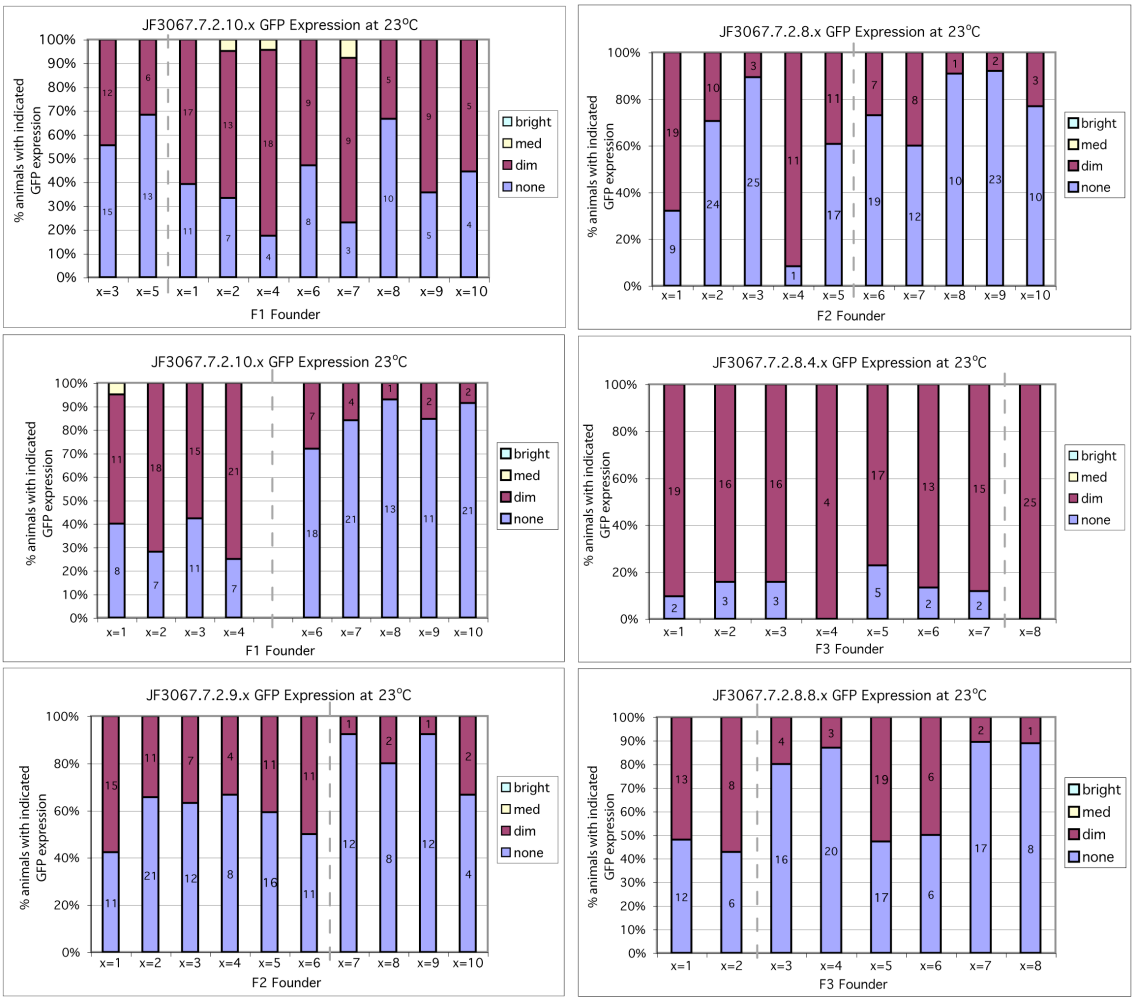


Figure 2.4. *ccEx3815* expression. (Top panel) Mosaic bodywall muscle GFP expression of an adult PD3815 animal. (Bottom panel) Mixed stage population of PD3815 viewed under a dissecting microscope. Animals were alive at observation.

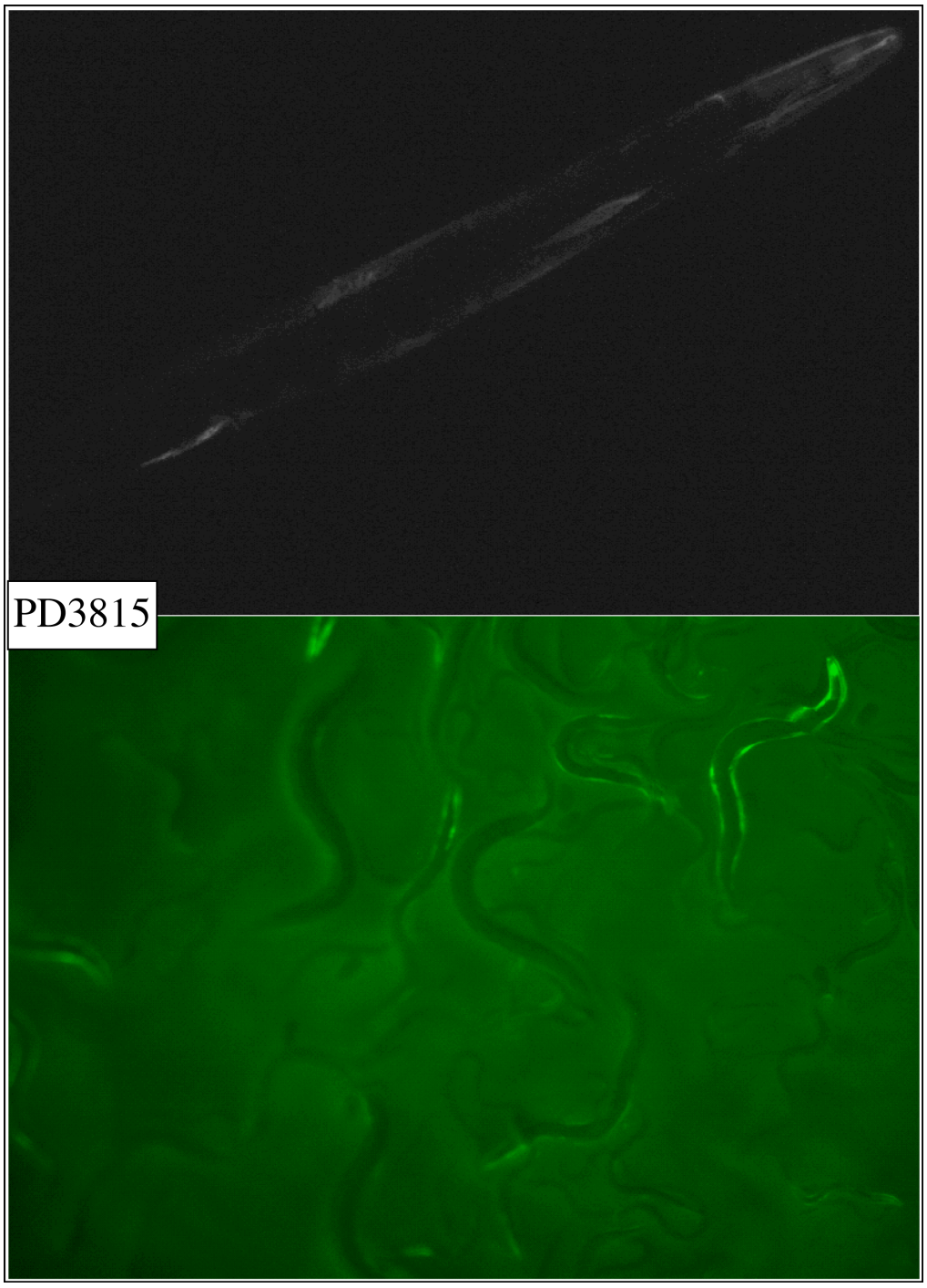


Figure 2.5. Logic behind the silencing screen. Endogenous *C. elegans* factors presumably act on a reporter gene to suppress its expression. Animals carrying such a silenced transgene exhibit weak and mosaic expression. If these endogenous factors are rendered non-functional by a mutagen, the reporter transgene would be de-repressed. Mutant animals would then show stronger, more uniform transgene expression.

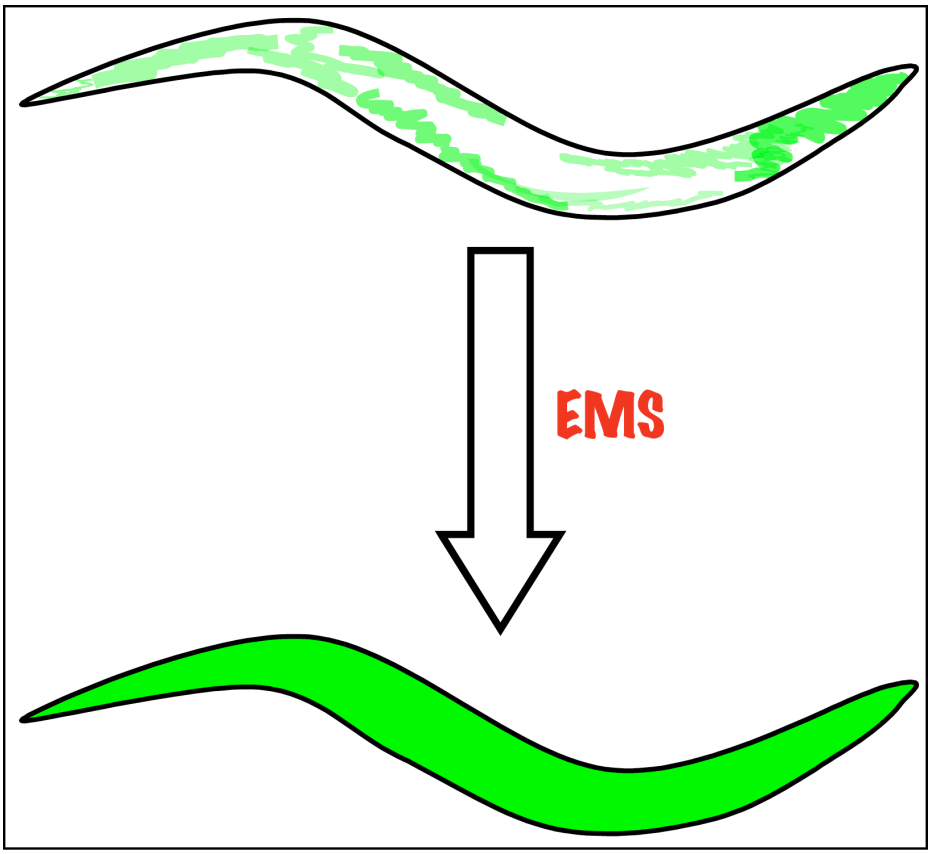


Figure 2.6. Diagrammatic representation of the screening procedure. The inset shows the genetics behind the silencing screen. The mutagenized chromosome is represented next to the asterisk. "*A*" is the extra-chromosomal array *ccEx3815*. Selfing is indicated by the symbol ⊗.

Figure 2.6

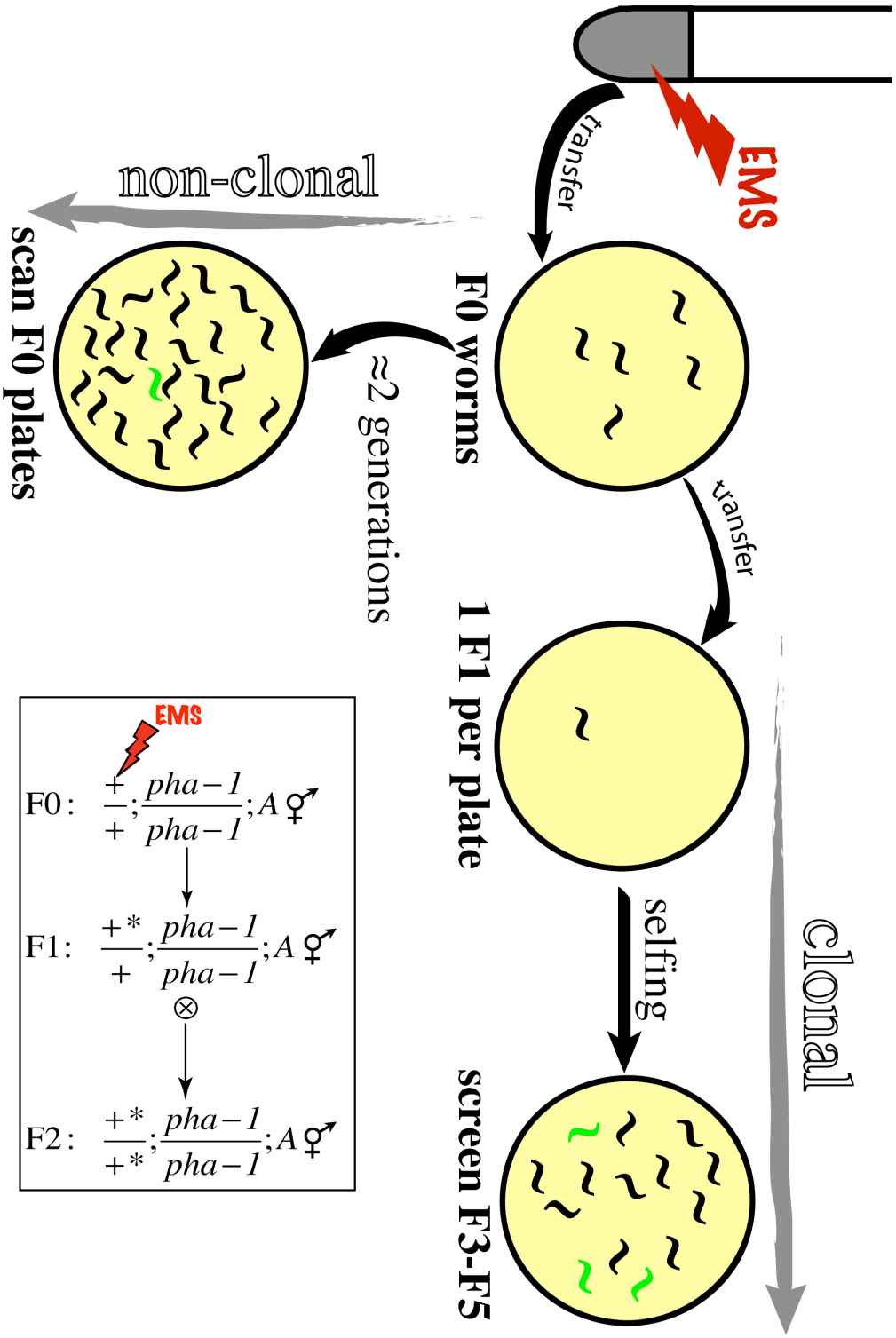

EMS
transfer
non-clonal
F0 worms
scan F0 plates
≈2 generations
transfer
1 F1 per plate
selfing
clonal
screen F3-F5
EMS
F0 : $\frac{+}{+};\frac{pha-1}{pha-1};A\,♂$
F1 : $\frac{+*}{+};\frac{pha-1}{pha-1};A\,♂$
⊗
F2 : $\frac{+*}{+*};\frac{pha-1}{pha-1};A\,♂$

Figure 2.7. GFP expression of a candidate *bright* mutant. GFP expression of candidate mutant PD3861 compared to PD3815 under a compound microscope (top panels) and dissecting microscope (bottom panels). For compound microscopy, animals were immobilized in levamisole (see Materials and Methods). Animals were either alive or recently deceased at time of observation. For dissecting microscopy, animals were observed live.

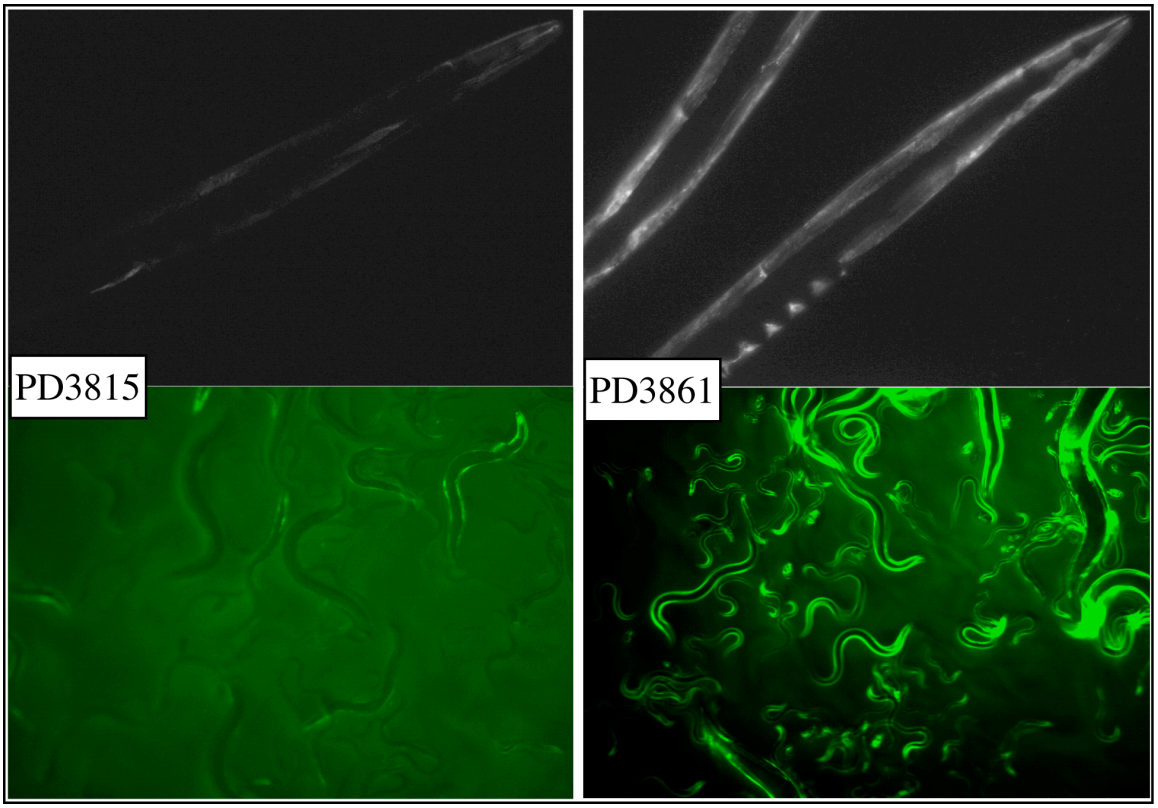
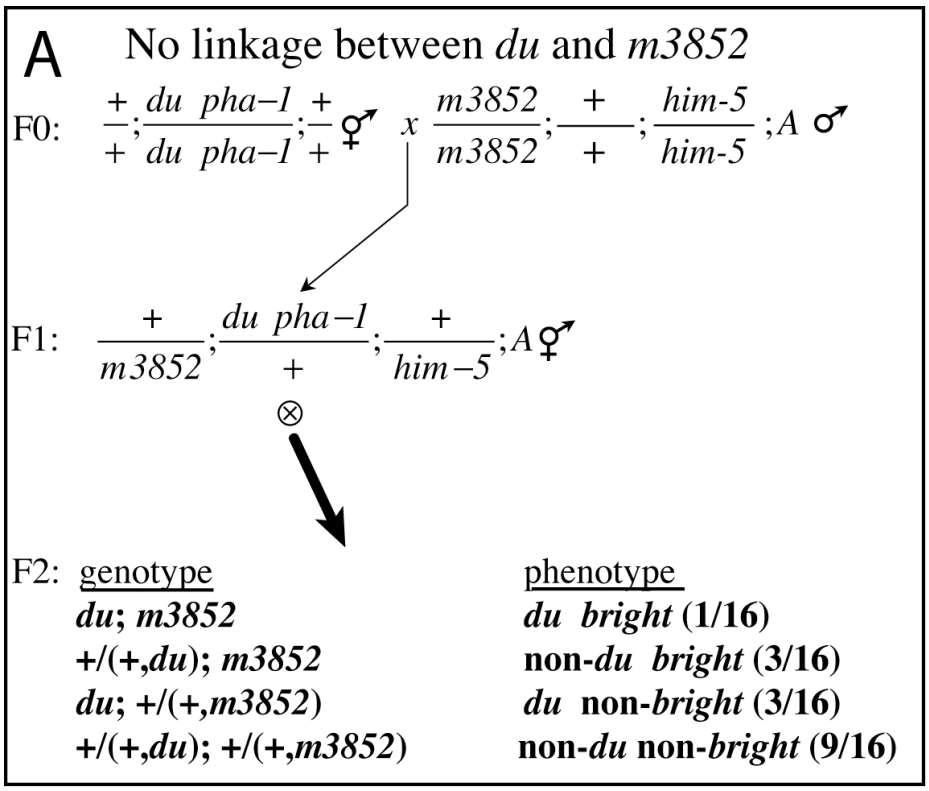


Figure 2.8. Mapping *bright* candidate PD3852. (A) Procedure and outcome of mapping the hypothetical recessive mutation, *m3852*, that is not linked to the marker chromosome. (B) Procedure and outcome of mapping the hypothetical recessive mutation that is linked to the marker chromosome. (C) Actual outcome of mapping the mutation in PD3852 to a chromosome. The inset shows the genetic interpretation of the mapping data.
du (*dpy unc*) are two generic linked recessive markers; "*A*" is the extra-chromosomal transgene array

Figure 2.8


A
No linkage between du and m3852
F0: +/+;du pha−1/du pha−1;+/+ ♀ x m3852/m3852;+/+;him-5/him-5;A ♂
F1: +/m3852;du pha−1/+;+/him−5;A ♀
⊗
F2: genotype
du; m3852
+/(+,du); m3852
du; +/(+,m3852)
+/(+,du); +/(+,m3852)
phenotype
du bright (1/16)
non-du bright (3/16)
du non-bright (3/16)
non-du non-bright (9/16)

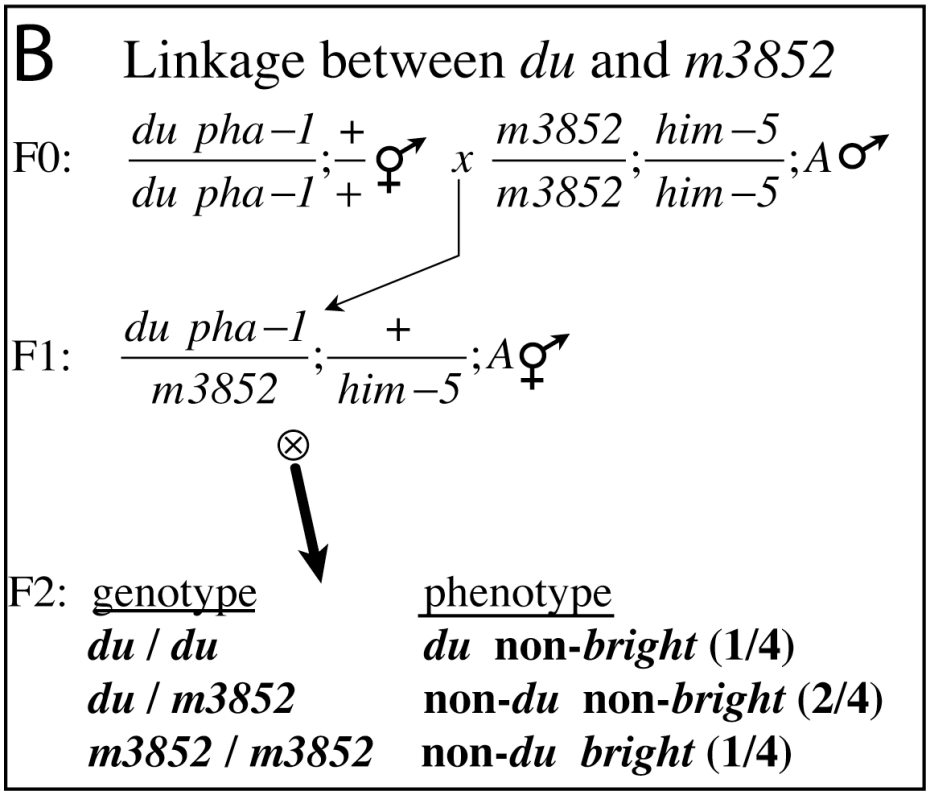

B
Linkage between du and m3852
F0: du pha−1/du pha−1;+/+ ♀ x m3852/m3852;him−5/him−5;A ♂
F1: du pha−1/m3852;+/him−5;A ♀
⊗
F2: genotype
du / du
du / m3852
m3852 / m3852
phenotype
du non-bright (1/4)
non-du non-bright (2/4)
non-du bright (1/4)

Figure 2.8 (*continued*)

C

F0: *pha-1 du* / *pha-1 du* ♂ x *pha-1* / *pha-1*; *m3852* / *m3852*; *A* ♂

F1: *pha-1 du* / *pha-1* ; + / *m3852* ; *A* ♀

⊗

F2: All F1's segregated the following F2 classes:

bright non-*du*

+

WT(non-*bright* non-*du*)

F2: 55 WT ♂

⊗

F3: 50/55 F2's segregated: non-*du bright* + WT

5/55 F2's segregated: non-*du bright* + WT + *dpy bright* + *dpy* non-*bright*

33 *dpy* non-*bright* ♂

⊗

F4: 33/33 segregated: *dpy bright* + *dpy* non-*bright*

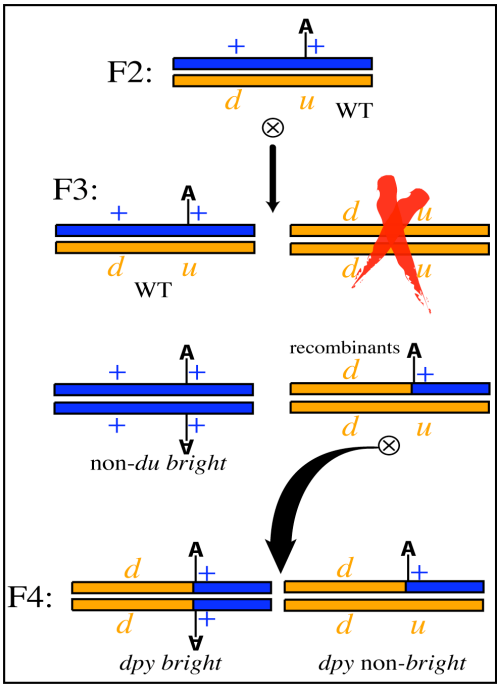


Figure 2.9. A genetic test for second-site mutations in PD3852. We make the assumption that the *bright* phenotype is conferred by a recessive mutation (*m*) plus at least one chromosome carrying the integrated array or by array homozygosity (regardless of the presence of *m*). (A) Since there is no way to distinguish between these two possibilities, all *bright* F2 worms were picked into 16°C and allowed to self. (B) Only array hemizygotes can produce array-minus progeny at 16°C. Note that A and A' are distinct arrays. A' is the extra-chromosomal naïve array from PD3815 that has not "seen" the PD3852 background.

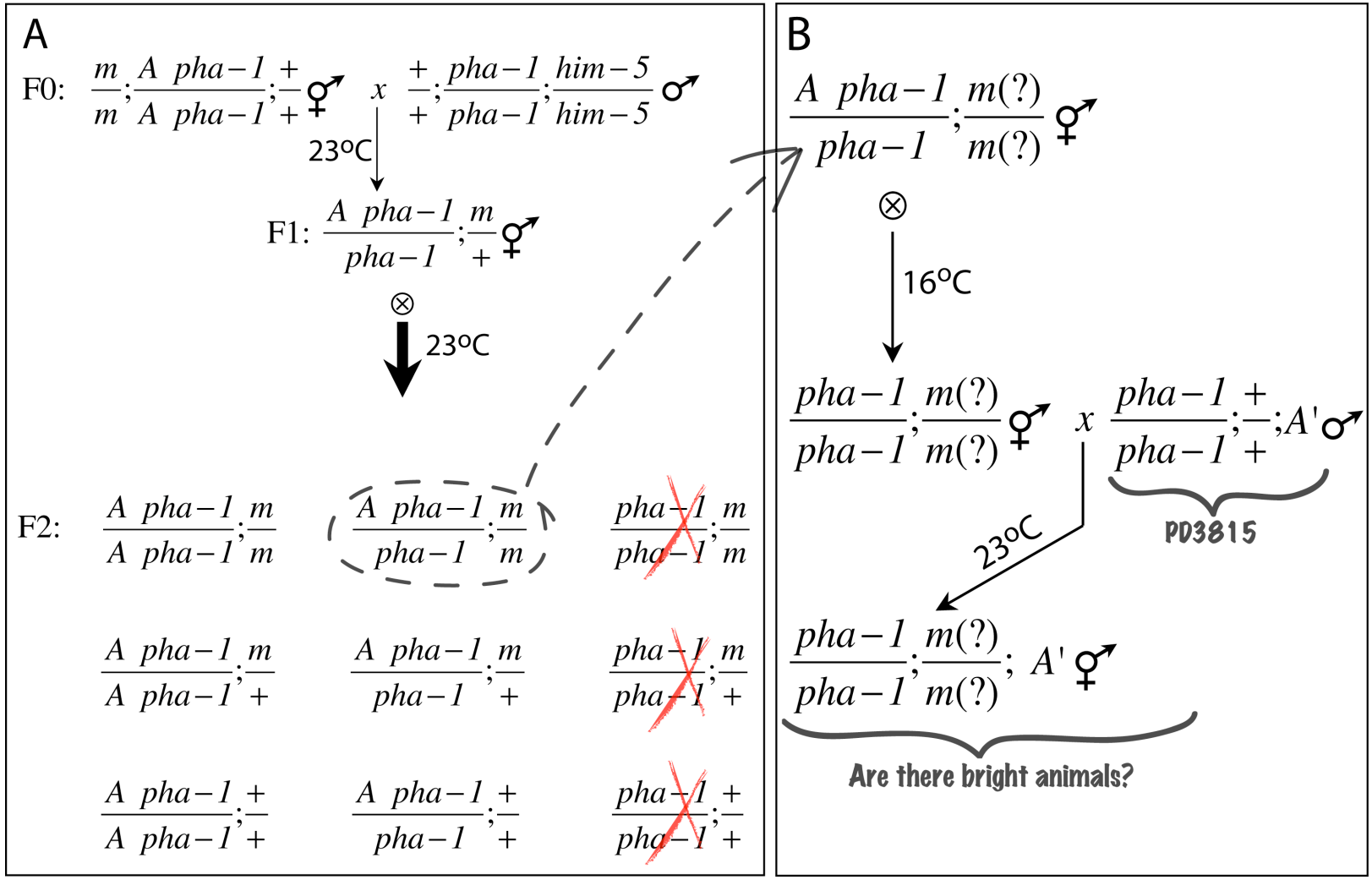
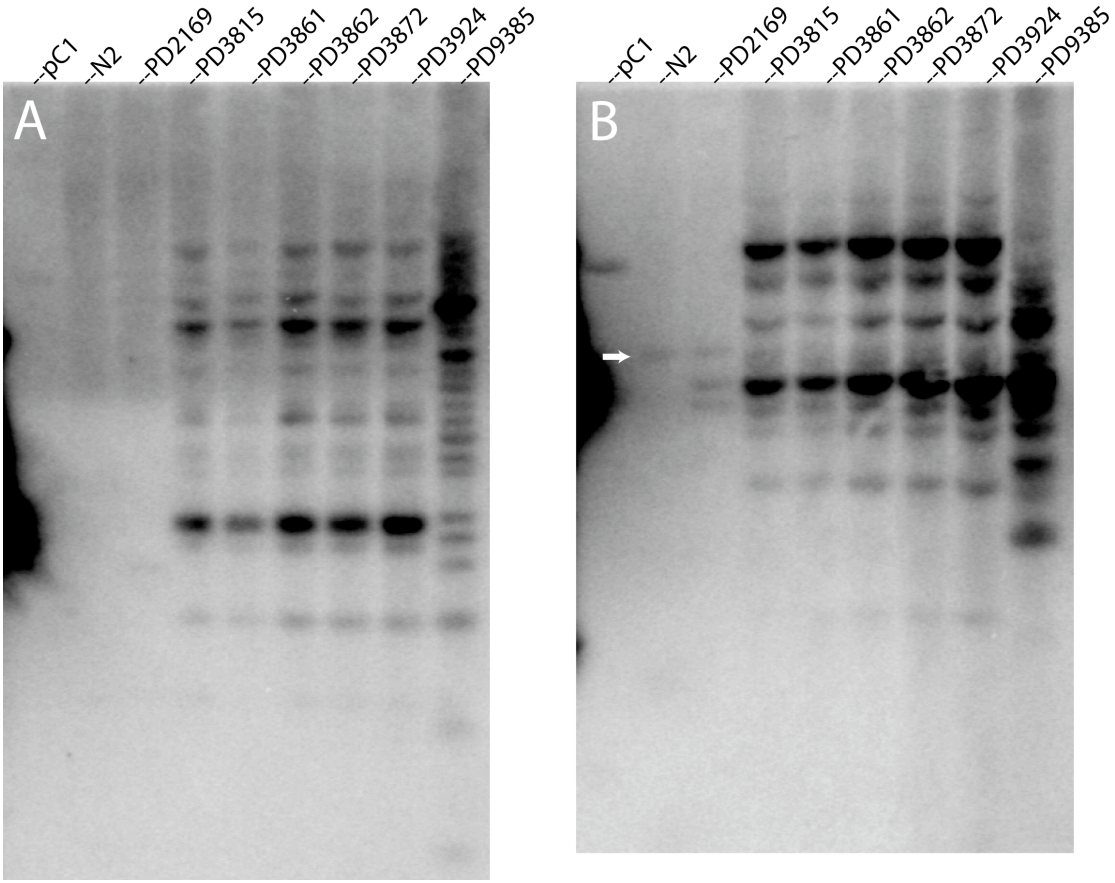


Figure 2.9

Figure 2.10. Southern hybridizations of *ccEx3815* and its derivatives. (A) Southern blot from a *Pst I* plus *PspOM I* double digest of genomic DNA. The probe is a segment from the *unc-54* promoter. There is a weak band in the pC1 lane [*pha-1*(*2123ts*) rescue construct], possibly due to weak homology between the probe and pC1. The N2 *unc-54* band is not visible in this blot. (B) Southern blot from a *Mfe I* plus *PspOM I* double digest of genomic DNA. The same probe as Southern blot A was used. The N2 *unc-54* band is indicated by the arrow. The faint PD2169 bands are possibly due to weak sequence homology between *edIs6* and the probe. Signals from both Panels A and B are not quantitative. PD3872 is PD3852 outcrossed once. PD3924 is PD3862 marked *in cis* with *dpy-5*(*e61*) and *unc-13*(*e1091*). PD2169 carries the integrated transgene *edIs6*[*unc-119::gfp*]. PD9385 carries an integrated *unc-54::gfp* array not derived from *ccEx3815*. The probe to both Southern blots is a segment from the *unc-54* promoter.

Figure 2.10


pC1
N2
PD2169
PD3815
PD3861
PD3862
PD3872
PD3924
PD9385
A
pC1
N2
PD2169
PD3815
PD3861
PD3862
PD3872
PD3924
PD9385
B

TABLES

Table 2.1. Summary of *bright* candidates from EMS screen. This table lists the five strongest *bright* candidates, including their penetrance and expressivity, and other phenotypes if present.

Table 2.1: *bright* candidates from first EMS screen		
candidate	**Phenotype**	**penetrance**
PD3852	• *bright*(+++) • transgene array integrated into LG III • lagging *unc*	100%
PD3861	• *bright*(++) • transgene array integrated into LG V	100%
PD3862	• *bright*(++) • transgene array integrated into LG I	100%
PD3870	• *bright*(++) • transgene array integrated • slightly *dpy*	100%
PD3891	• *bright*(+) • transgene array probably integrated • sterile at 16°C • slightly *unc*(?)	≈90%

ACKNOWLEDGMENTS

We acknowledge Jamie Fleenor for performing the microinjection procedure.

REFERENCES

AMEDEO, P., Y. HABU, K. AFSAR, O. MITTELSTEN SCHEID and J. PASZKOWSKI, 2000 Disruption of the plant gene MOM releases transcriptional silencing of methylated genes. Nature **405:** 203-206.

ANDERSON, P., 1995 Mutagenesis, pp. 31-54 in *Caenorhabditis elegans: Modern Biological Analysis of an Organism*, edited by H. F. EPSTEIN and D. C. SHAKES. Academic Press, San Diego, CA.

BARNES, T. M., Y. KOHARA, A. COULSON and S. HEKIMI, 1995 Meiotic recombination, noncoding DNA and genomic organization in *Caenorhabditis elegans*. Genetics **141:** 159-179.

BRENNER, S., 1974 The genetics of *Caenorhabditis elegans*. Genetics **77:** 71-94.

CAPELSON, M., and V. G. CORCES, 2004 Boundary elements and nuclear organization. Biol Cell **96:** 617-629.

CONSORTIUM, 1998 Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science **282:** 2012-2018.

COSGROVE, M. S., and C. WOLBERGER, 2005 How does the histone code work? Biochem Cell Biol **83:** 468-476.

FANTI, L., D. R. DORER, M. BERLOCO, S. HENIKOFF and S. PIMPINELLI, 1998 Heterochromatin protein 1 binds transgene arrays. Chromosoma **107:** 286-292.

FIRE, A., D. ALBERTSON, S. W. HARRISON and D. G. MOERMAN, 1991 Production of antisense RNA leads to effective and specific inhibition of gene expression in *C. elegans* muscle. Development **113:** 503-514.

FIRE, A., and C. MELLO, 1995 Transformation, pp. 451-482 in *Caenorhabditis elegans: Modern Biological Analysis of an Organism*, edited by H. F. EPSTEIN and D. C. SHAKES. Academic Press, San Diego.

GRANATO, M., H. SCHNABEL and R. SCHNABEL, 1994 *pha-1*, a selectable marker for gene transfer in *C. elegans*. Nucleic Acids Res **22:** 1762-1763.

GREENWALD, I., A. COULSON, J. SULSTON and J. PRIESS, 1987 Correlation of the physical and genetic maps in the lin-12 region of *Caenorhabditis elegans*. Nucleic Acids Res **15:** 2295-2307.

GRIFFITHS, A. J. F., J. H. MILLER, D. T. SUZUKI, R. C. LEWONTIN and W. M. GELBART, 2000 Mechanisms of Gene Mutation, pp. 495-522 in *An Introduction to Genetic Analysis*. W. H. Freeman and Company, New York.

GRISHOK, A., H. TABARA and C. C. MELLO, 2000 Genetic requirements for inheritance of RNAi in C. elegans. Science **287:** 2494-2497.

HIROCHIKA, H., H. OKAMOTO and T. KAKUTANI, 2000 Silencing of retrotransposons in arabidopsis and reactivation by the ddm1 mutation. Plant Cell **12:** 357-369.

HSIEH, J., J. LIU, S. A. KOSTAS, C. CHANG, P. W. STERNBERG *et al.*, 1999 The RING finger/B-box factor TAM-1 and a retinoblastoma-like protein LIN-35 modulate context-dependent gene silencing in *Caenorhabditis elegans*. Genes Dev **13:** 2958-2970.

HSIEH, J., and A. FIRE, 2000 Recognition and silencing of repeated DNA. Annu Rev Genet **34:** 187-204.

JANSEN, G., E. HAZENDONK, K. L. THIJSSEN and R. H. PLASTERK, 1997 Reverse genetics by chemical mutagenesis in *Caenorhabditis elegans*. Nat Genet **17:** 119-121.

JEDDELOH, J. A., T. L. STOKES and E. J. RICHARDS, 1999 Maintenance of genomic methylation requires a SWI2/SNF2-like protein. Nat Genet **22:** 94-97.

KETTING, R. F., T. H. HAVERKAMP, H. G. VAN LUENEN and R. H. PLASTERK, 1999 Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell **99:** 133-141.

LABRADOR, M., and V. G. CORCES, 2002 Setting the boundaries of chromatin domains and nuclear organization. Cell **111:** 151-154.

MADURO, M., and D. PILGRIM, 1995 Identification and cloning of *unc-119*, a gene expressed in the *Caenorhabditis elegans* nervous system. Genetics **141:** 977-988.

MARTIN, C., and Y. ZHANG, 2005 The diverse functions of histone lysine methylation. Nat Rev Mol Cell Biol **6:** 838-849.

MELLO, C. C., J. M. KRAMER, D. STINCHCOMB and V. AMBROS, 1991 Efficient gene transfer in *C.elegans*: extrachromosomal maintenance and integration of transforming sequences. Embo J **10:** 3959-3970.

PRASAD, S. S., and D. L. BAILLIE, 1989 Evolutionarily conserved coding sequences in the dpy-20-unc-22 region of *Caenorhabditis elegans*. Genomics **5:** 185-198.

SCHNABEL, H., and R. SCHNABEL, 1990 An organ-specific differentiation gene, *pha-1*, from *Caenorhabditis elegans*. Science **250:** 686-688.

STARR, T., A. M. HOWELL, J. MCDOWALL, K. PETERS and A. M. ROSE, 1989 Isolation and mapping of DNA probes within the linkage group I gene cluster of *Caenorhabditis elegans*. Genome **32:** 365-372.

STINCHCOMB, D. T., J. E. SHAW, S. H. CARR and D. HIRSH, 1985 Extrachromosomal DNA transformation of *Caenorhabditis elegans*. Mol Cell Biol **5:** 3484-3496.

TAMARU, H., and E. U. SELKER, 2001 A histone H3 methyltransferase controls DNA methylation in Neurospora crassa. Nature **414:** 277-283.

VASTENHOUW, N. L., S. E. FISCHER, V. J. ROBERT, K. L. THIJSSEN, A. G. FRASER *et al.*, 2003 A genome-wide screen identifies 27 genes involved in transposon silencing in C. elegans. Curr Biol **13:** 1311-1316.

WOOD, A., J. SCHNEIDER and A. SHILATIFARD, 2005 Cross-talking histones: implications for the regulation of gene expression and DNA repair. Biochem Cell Biol **83:** 460-467.

WOODHOUSE, M. R., M. FREELING and D. LISCH, 2006 The mop1 (mediator of paramutation1) mutant progressively reactivates one of the two genes encoded by the MuDR transposon in maize. Genetics **172:** 579-592.

CHAPTER 2

Part II: Forward genetic screen for *C. elegans* silencing mutants

INTRODUCTION

In Part I of Chapter 2, we described a genetic screen to isolate silencing mutants in *C. elegans*. We established a transgenic line carrying the extra-chromosomal array *ccEx3815*. This array contains multiples copies of the *unc-54::gfp* construct, possibly arranged in tandem repeats. *ccEx3815* exhibits weak and mosaic expression in transgenic animals. Our screen led to the isolation of a handful of candidates that exhibited strong, uniform GFP expression through the bodywall muscles. Further characterization of three candidates led to the realization that second-site mutations did not exist in these three candidate mutants. For reasons we do not yet understand, integration of *ccEx3815* into *C. elegans* chromosomes de-silenced the integrated derivatives without any apparent change in the structure of the array. This was true for at least three integrated derivates of *ccEx3815* (*ccIn3852*, *ccIn3861*, and *ccIn3861*). We did not antipicate that EMS mutagenesis would lead to high frequencies of transgene integration, as the method of choice for transgene integration in *C. elegans* is irradiation. Thus, it appears that we unintentionally selected for integration events in our screen.

We designed a different strategy to screen for silencing mutants. Our strategy was to prevent selection of integration events. Whereas in Part I we mutagenized hermaphrodite animals harboring *ccEx3815*, in our new strategy we mutagenized males and introduced *ccEx3815* into mutagenized genomes via mating.

We isolated about a dozen candidates from the second screen. Further characterization of one candidate indicated that (1) the transgene remained extra-chromosomal, and (2) the phenotype was conferred by the genome. However, we were not able to identify linkage groups for this (and other) candidates.

MATERIALS AND METHODS

Materials and methodology used were as described in Part I. The following additional transgenic lines were used in the experiments described in Part II:

PD3822 [*pha-1*(*e2123ts*) *dpy-17*(*e164*) *unc-32*(*e189*) III; *ccEx3815*]: *pha-1*(*e2123ts*) marked *in cis* with the two recessive markers *dpy-17*(*e164*) and *unc-32*(*e189*) and carrying the extra-chromosomal transgene *ccEx3815*

PD3838 [*pha-1*(*e2123ts*) III; *ccEx3815*]: putative *bright* mutant

RESULTS

Isolation of putative silencing mutants

We carried out six EMS mutageneses and screened for silencing mutants according to the diagram in Figure 2.11. Briefly, mutagenized males were crossed into a recessively marked line carrying *ccEx3815*. F1 progeny were transferred individually, each to a plate, and allowed to produce progeny. Because crosses normally result in both male and hermaphrodite progeny, to avoid sib-mating, we took care to use only F1 hermaphrodites at the L4 larvae stage. A clonal screen of 5,784 F1's plus a non-clonal screen of ≈240 F1's (two F1's per plate) yielded seven candidates plus 10 putative mutants with weak phenotypes (Table 2.2). Each mutagenesis resulted in 3.5-11.4% F1 lethality or sterility. This is within the dosage range for a typical EMS mutagenesis. We chose candidate PD3838 for further characterization.

PD3838 exhibits nebulous properties

Figure 2.12 shows a PD3838 adult animal. Compared to PD3815, GFP expression in PD3838 is stronger and more uniform. However, a population of PD3838 as a whole displays some heterogeneity, with varying degrees of expression and mosaicism among animals. In any given generation derived from a single founder, the percent of *bright* animals ranges from 30-90% (single generation penetrance). In mixed generation populations, the penetrance is about 50-60%. From extensive pedigree analyses of the line (below), we can make the generalization that a *bright* PD3838 animal produces at least 20% *bright* progeny; while non-*bright* animals rarely achieve this percentage. Qualitatively, a population derived from a *bright* founder is usually unequivocal for the phenotype (level of expression and penetrance). A non-*bright* founder usually produces a

population that is rather vague and quite difficult to judge. Thus, there appears to be some level of noise inherit in PD3838, manifested in populations with low penetrance. From extensive work with the strain, we estimate the level of background noise to be about 20%.

Anecdotal evidence suggests that the activity of the transgene array in PD3838 might be buffered against sudden changes in expression state. For example, when PD3838 is initially brought out of starvation, the *bright* phenotype is not apparent, and the line appears similar to PD3815 in GFP expression level. Passage of several generations is required before the *bright* phenotype emerges from the fed population. During the course of stock passage of PD3838, we have observed that the number of *bright* animals in the (continuously fed) population seem to increase over time.

Pedigree analysis reveals the putative locus to be potentially dominant

To gain further insight into the nebulous nature of PD3838, we performed a pedigree analysis. Individual *bright* or non-*bright* animals (F0) were allowed to self and the immediate next generation (F1) scored for percent that were *bright*. The result of this analysis is shown in Figure 2.13. It is evident in Figure 2.13A that non-*bright* parents (#1-10) give rise to non-*bright* populations while *bright* parents (#11-20) give rise to predominantly *bright* populations. The experiment was reiterated for populations 6.x, 12.x, and 17.x with the same result (Figure 2.13B). Founders 12.6 and 12.8 may have been non-expressing mutants. Thus, the pedigree analysis reveals a segregation pattern consistent with a dominant locus: non-*bright* parents (+/+) produce only non-*bright* progeny; whereas *bright* parents (m^D/+) give rise to both *bright* and non-*bright* progeny. This is supported by preliminary mapping data indicating a putative dominant locus on

Chromosome II (data not shown). Interestingly, preliminary pedigree analyses of the other candidates shown in Table 2.2 indicate that they all may be segregating as dominant loci.

The *bright* phenotype can be recovered by introduction of a naïve array into the PD3838 background

We sought to determine whether the transgene array in PD3838 had integrated into the genome. Assuming a dominant locus, we performed the experiment shown in Figure 2.14A. A single PD3838 *bright* animal was shifted to 16°C and allowed to produce self progeny. At 16°C, some F1 animals are born without the transgene array. These animals are crossed singly to males carrying a naïve array from PD3815 (i.e. *ccEx3815*, represented as *A'* in the figure) and reared at 23°C for *pha-1*(*e2123ts*) selection. The resulting F2 animals are pooled (blindly without knowing their GFP status) a few animals per plate. In F2 pools that segregate *bright* progeny, a few *bright* animals are again pooled to see if they produce a population enriched in *bright* animals. If so, *bright* animals are selfed and the percent of *bright* progeny determined.

Figure 2.14B shows the pedigree resulting from the analysis. Two generalizations can be made from the data. First, although the trend is not so clear cut, one can see that introduction of the naïve array requires passage for at least two generations before the array becomes active in the majority of animals. (This is consistent with our observation above that the transgene array is not immediately active after the strain is initially brought out of starvation). This is apparent in lineages 12.p6.2.x and 19.p7.2.x. In these two lineages, we have recovered the *bright* phenotype only after the naïve array has "seen" the PD3838 background for approximately four generations. A control experiment in

which the same naïve array was introduced into a wildtype background [*pha-1*(*e2123ts*) animals], did not recover the *bright* phenotype (data not shown). Second, as mentioned previously, there appears to be some level of noise up to the 20% level. Lineages with less than 20% single-generation penetrance do not appear to breed true.

We extended the pedigree analysis further by determining whether we could enrich for *bright* animals in a population founded by a single non-*bright* progenitor. A non-*bright* L4 animal from lineage 18.p8 (Figure 2.14B) was allowed to produce self progeny. The fraction of *bright* (multi-generation) progeny resulting from this animal was approximately 1-2%. Five non-*bright* and five *bright* progeny (designated F0 generation) were selfed and F1 progeny scored for expression. We then selected the brightest F1 animal from each group to be the founder of the F2 generation and scored the F2 for expression. The procedure was reiterated with the brightest F2 from each group, and so on until five generations of *bright* selection were obtained. The result of the analysis is shown in Figure 2.15. In column A, the F0 founders were all non-*bright* while in column B the F0 founders were all *bright*. In some cases in column A, no *bright* animals existed in the population and a non-*bright* animal was used instead. The analysis shows that to a small extent, we can enrich for *bright* animals from a single non-*bright* founder. However, we can never obtain the high percentage of *bright* animals from a non-*bright* founder than we can from a *bright* founder (compare Figure 2.13 to Figure 2.15).

PD3838 does not appear to be a mutator strain

We wondered if PD3838 might be a mutator strain. During the course of our work with the strain we encountered the appearance of other phenotypes (mostly *dpy* and/or *unc* but also some *egl* and blunt-tailed animals) emerging at frequencies higher than in

wildtype populations. Additionally, during the course of pedigree analysis experiments, certain lineages were found to be 100% lethal or sterile (Figure 2.14B, lineages 21 and 23). If PD3838 were a mutator strain, then transposon activities would cause insertional mutagenesis, leading to loss-of-function of target genes, and we should be able to recover lines carrying mutations unrelated to PD3838. Upon encountering a morphologically defective animal in the population, we selfed the animal to see if it bred true for the phenotype. None of the animals (about 25 total) bred true. Thus, PD3838 is not a mutator strain. However, if the transposon activity only affected the soma (or affected the soma at a higher frequency than the germline), then unrelated mutations would not be recovered from PD3838.

DISCUSSION

We attempted an EMS screen for *C. elegans trans*-acting factors whose loss-of-function or gain-of-function mutation would lead to de-repression of a silenced reporter transgene, *ccEx3815*. We isolated several candidates including one, PD3838, that we chose to characterize in detail. The phenotype of this line is rather intractable. The phenotype is not completely penetrant, with a high background noise of up to 20%. The transgene array appears to require several generations to turn on when animals are brought out of starvation. The locus appears to be metastable. Our initial effort to map the locus failed to identify any linkage. However, a second mapping attempt (performed approximately two months later), tentatively put the linkage to Chromosome II. An outcrossed derivative of PD3838 displays similar behavior (although we have not attempted to map the outcrossed strain).

In spite of the inherit difficulties in analyzing PD3838, we believe the *bright* phenotype is linked to the genome. First, the *bright* phenotype does not map to the transgene array, as the phenotype can be recovered by introduction of a naïve array into the PD3838 background. Second, pedigree analysis indicates that PD3838 behaves distinctly from PD3815. PD3838 breeds true for the *bright* phenotype (within the 20% error limit; Figure 2.13); a *bright* or non-*bright* lineage can be established and maintained (Figure 2.14B, lineages 12.p6.2.x and 19.p7.2.x). In contrast, in PD3815, we can only partially select for *bright* or non-*bright* character for only one generation (Chapter 2 Part I, Figure 2.3). Third, we are able to find a linkage for the locus, though only preliminarily.

Difficulty in finding a linkage group and frequent appearance of unrelated phenotypes in a population of PD3838 are two characteristics of mutator strains found in *C. elegans* (COLLINS *et al.* 1987; GRISHOK *et al.* 2000; KETTING *et al.* 1999; MOERMAN and WATERSTON 1984; POTHOF *et al.* 2003; VASTENHOUW *et al.* 2003). Previously silent transposons become active, causing insertional mutagenesis in the genome and a high frequency of new mutations to arise in the population. Although we have observed the frequent appearance of unrelated phenotypes during maintenance of the stock, these phenotypes do not breed true. Thus, PD3838 is not likely to be a mutator strain.

In other model systems there are well-documented cases of mutant loci that are metastable. Such "epimutations" or "epi-alleles" are meiotically heritable and often exhibit unstable phenotypes and non-Mendelian inheritance (RICHARDS 2006). Research into epimutations has been most mature in the plant field, where epigenetic phenomena such as paramutation and epi-alleles were among the first to be described (BENDER and FINK 1995; BOWMAN *et al.* 1992; BRINK 1958; CHANDLER *et al.* 2000; DAS and MESSING 1994; JACOBSEN and MEYEROWITZ 1997; MEYER *et al.* 1993; STAM *et al.* 2002). More recently, epi-alleles have also been described in mice (MORGAN *et al.* 1999; RAKYAN *et al.* 2003; RASSOULZADEGAN *et al.* 2006) and fungi (COLOT *et al.* 1996). To date, epimutations have not been reported for *C. elegans*, but the nature of PD3838 is reminiscent of metastable alleles documented in other systems. If this is the case, the transgene array may retain its activity from the previous generation even when it has segregated away from the mutant background. This would leave the array in the "on" state (perhaps for an initial one or two generations) in a wildtype background, eventually shutting off if it continues to be propagated in the wildtype background.

In conclusion, detailed analysis of a putative silencing mutant reveals the locus to be metastable, possibly resulting from an epimutation. PD3838, as well as other candidates listed in Table 2.2, appears to segregate as a dominant locus. As the phenotype of PD3838 is rather fickle, any future analysis of the strain must be done at the population level and in a quantitative manner, perhaps involving QTL.

FIGURES

Figure 2.11. EMS screen for silencing mutants. The extra-chromosomal transgene *ccEx3815* is represented as "*A*" in the figure. *du* = *dpy-17*(*e164*) *unc-32*(*e189*) III

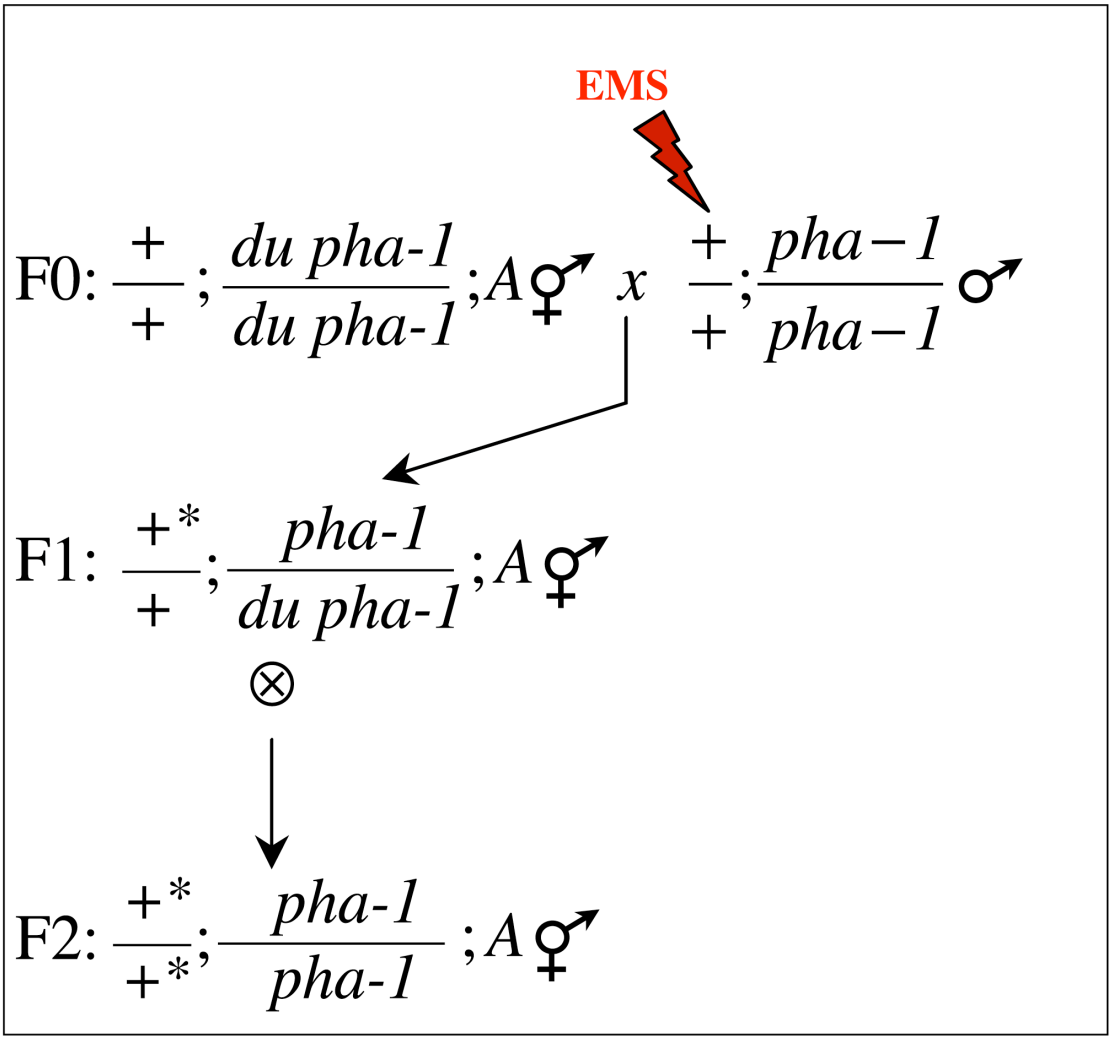


Figure 2.12. Candidate PD3838. Images of PD3815 and PD3838 taken with a dissecting microscope under identical settings. Animals were alive at time of observation.

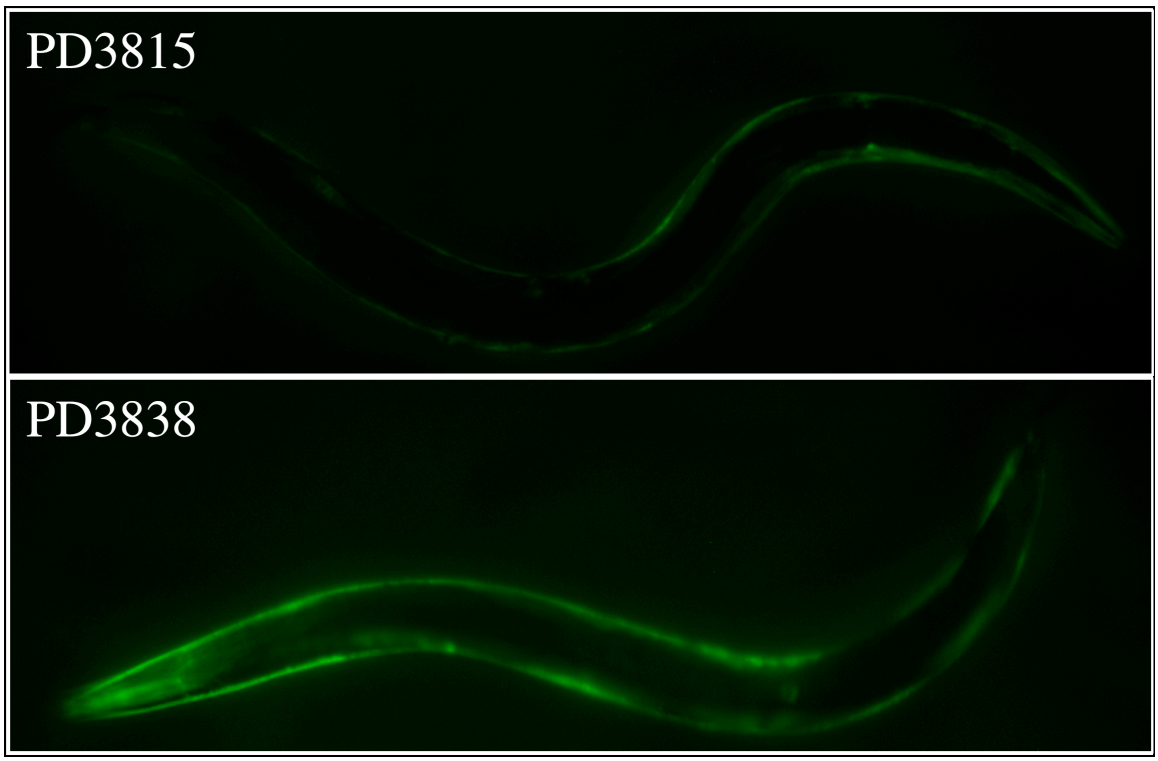


Figure 2.13. Pedigree analysis of PD3838. (A) The X-axis shows the founder (F0) animal. Non-*bright* founders are indicated with black and *bright* founders are indicated with green. Stacked vertical bars show the fraction of progeny (F1) that are *bright* (green) and non-*bright* (black). (B) The experiment is reiterated with the F1 progeny of founders 6, 12, and 17. Numbers in each column indicate the number of animals in that category.

Figure 2.13

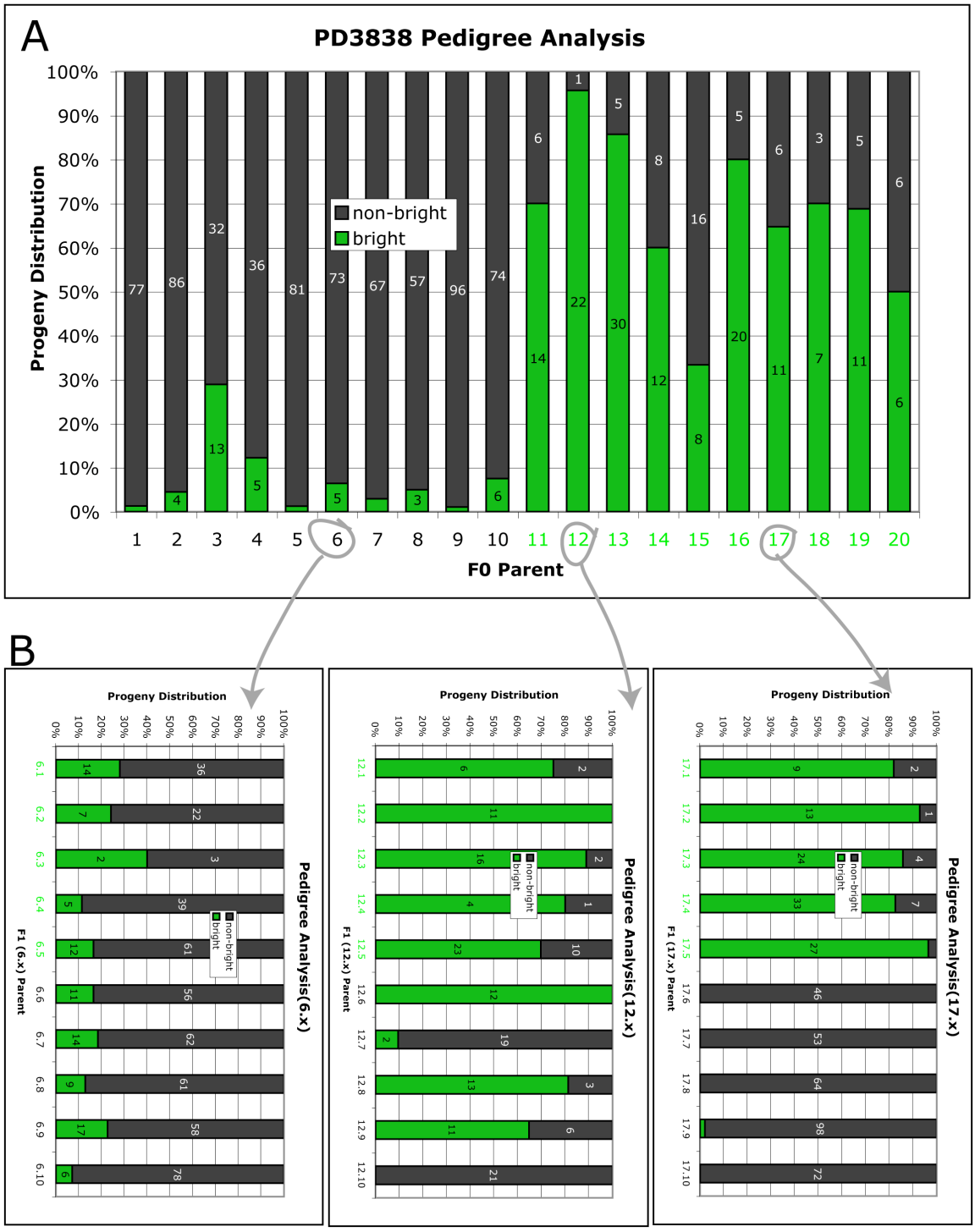

A
PD3838 Pedigree Analysis
Progeny Distribution
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
non-bright
bright
77 86 32 36 81 73 67 57 96 74 6 1 5 8 16 5 6 3 5 6
4 13 5 5 3 6 14 22 30 12 8 20 11 7 11 6
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
F0 Parent
B
Progeny Distribution
Pedigree Analysis(6.x)
F1 (6.x) Parent
6.1 14 36
6.2 7 22
6.3 2 3
6.4 5 39
6.5 12 61
6.6 11 56
6.7 14 62
6.8 9 61
6.9 17 58
6.10 6 78
Progeny Distribution
Pedigree Analysis(12.x)
F1 (12.x) Parent
12.1 6 2
12.2 11
12.3 16 2
12.4 4 1
12.5 23 10
12.6 12
12.7 2 19
12.8 13 3
12.9 11 6
12.10 21
Progeny Distribution
Pedigree Analysis(17.x)
F1 (17.x) Parent
17.1 9 2
17.2 13 1
17.3 24 4
17.4 33 7
17.5 27
17.6 46
17.7 53
17.8 64
17.9 98
17.10 72

Figure 2.14. Recovery of *bright* phenotype after introduction of a naïve array into PD3838. (A) Schematic representation of the experimental procedure. Note that array *A* and array *A'* both refer to *ccEx3815*. Array *A* was crossed into mutagenized animals, used in the screen (Figure 2.11), and propagated in PD3838. Array *A'* is the naïve array introduced into the PD3838 background for the first time. The dominant locus is represented as m^D. (B) Pedigree analysis of lineages resulting from the experiment depicted in (A). Each lineage is founded by a single F1 (top arrow). The F2 animals are pooled a few worms per plate (p1, p2, etc.). F2 pools that segregate *bright* animals are indicated in green. From each F2 pool that segregates *bright* animals, several *bright* animals are selfed. If the resulting (multi-generation) population segregates at least 20% *bright* animals (i.e. 12.p6.2), that population is indicated in green. If it segregates less than 20% *bright* animals, that population is indicated in black (i.e. 12.p6.1 and 12.p6.3). Note that all populations derived from the pooled plates (i.e. 12.p6.1 and 12.p6.3) contain some *bright* animals (because they were derived from *bright* founders). However, most do not have greater than 20% *bright* animals in the population. If a population has significantly more than the 20% cutoff of *bright* animals, the percentage of *bright* animals in the population is indicated (i.e. 12.p6.2 has greater than 50% *bright* animals in the population). In lineages marked with an "X", its founder(s) was either sterile or dead.

Figure 2.14

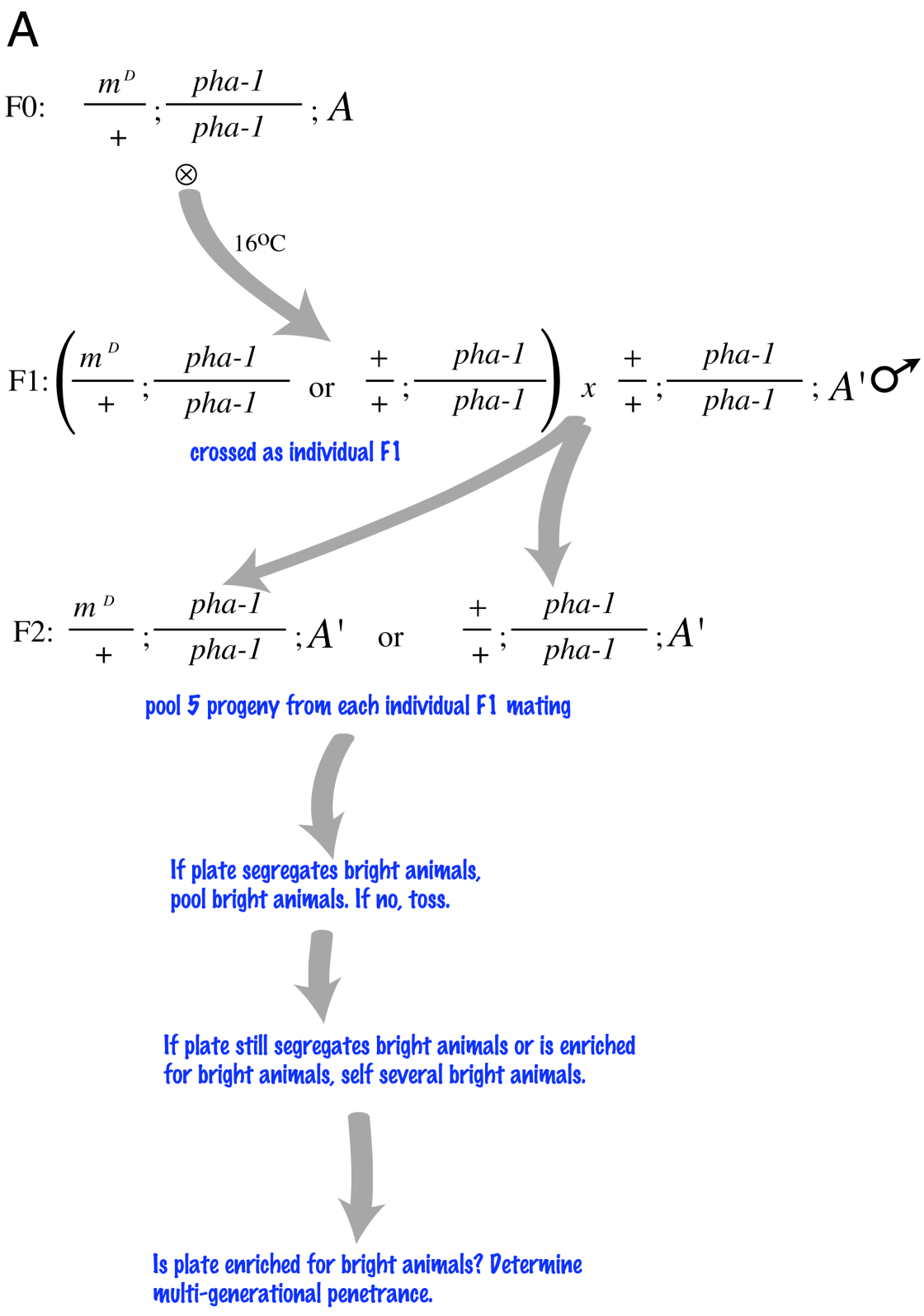

A
F0: $\frac{m^D}{+};\frac{pha\text{-}1}{pha\text{-}1};A$
⊗
16°C
F1: $\left(\frac{m^D}{+};\frac{pha\text{-}1}{pha\text{-}1}\ \text{or}\ \frac{+}{+};\frac{pha\text{-}1}{pha\text{-}1}\right)\ x\ \frac{+}{+};\frac{pha\text{-}1}{pha\text{-}1};A'♂$
crossed as individual F1
F2: $\frac{m^D}{+};\frac{pha\text{-}1}{pha\text{-}1};A'$ or $\frac{+}{+};\frac{pha\text{-}1}{pha\text{-}1};A'$
pool 5 progeny from each individual F1 mating
If plate segregates bright animals, pool bright animals. If no, toss.
If plate still segregates bright animals or is enriched for bright animals, self several bright animals.
Is plate enriched for bright animals? Determine multi-generational penetrance.

Figure 2.14 (*continued*)

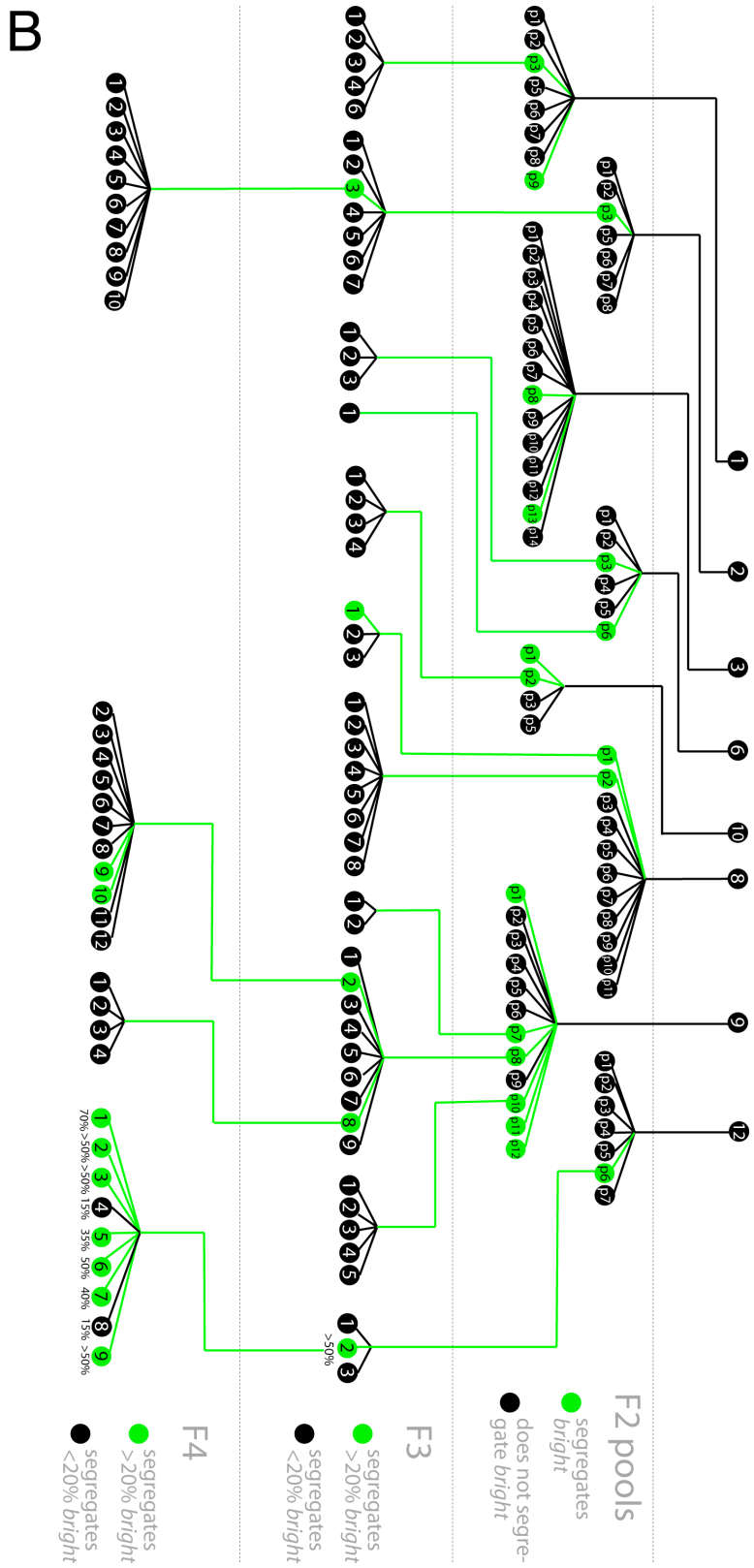


Figure 2.14 (*continued*)

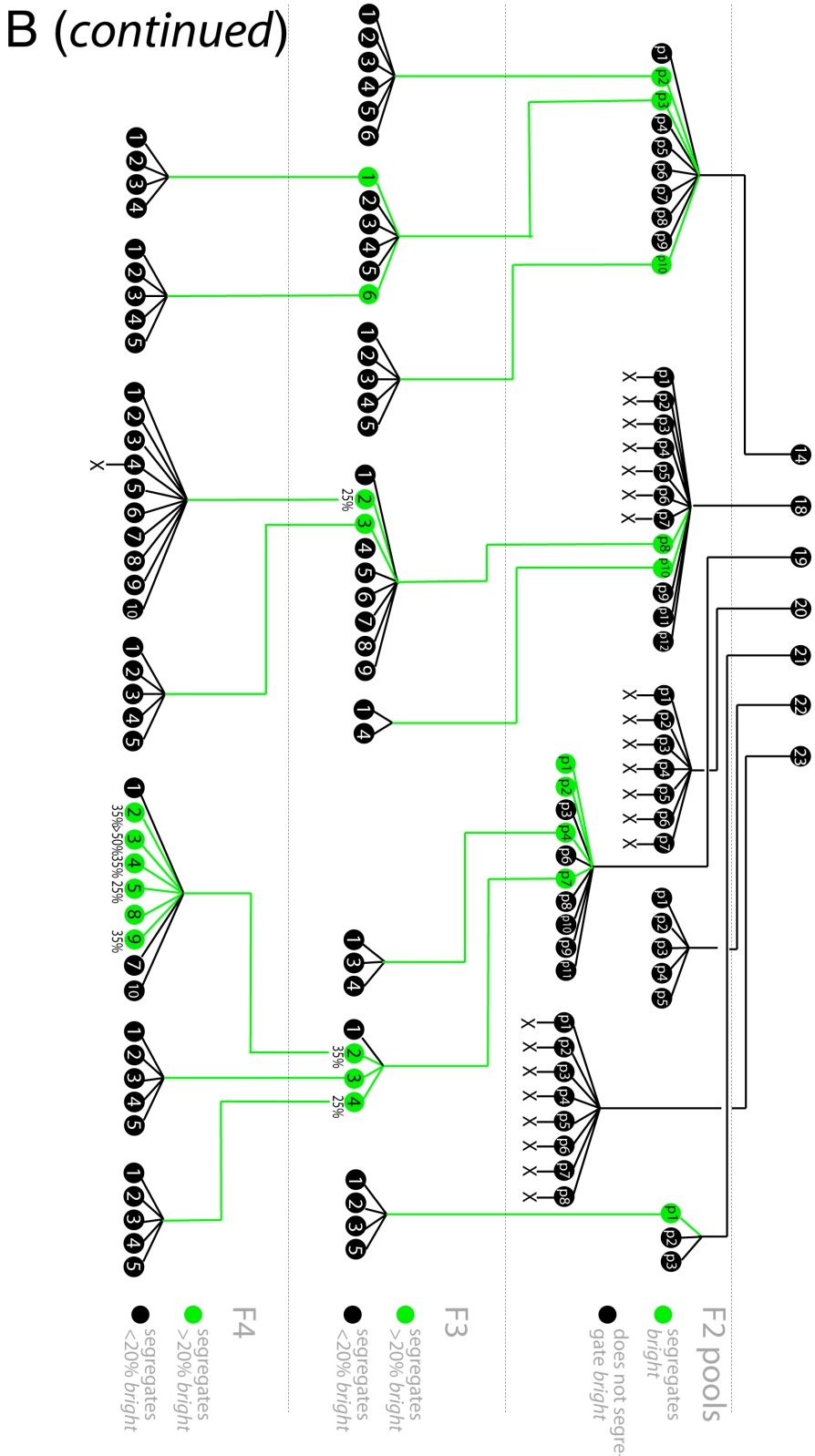
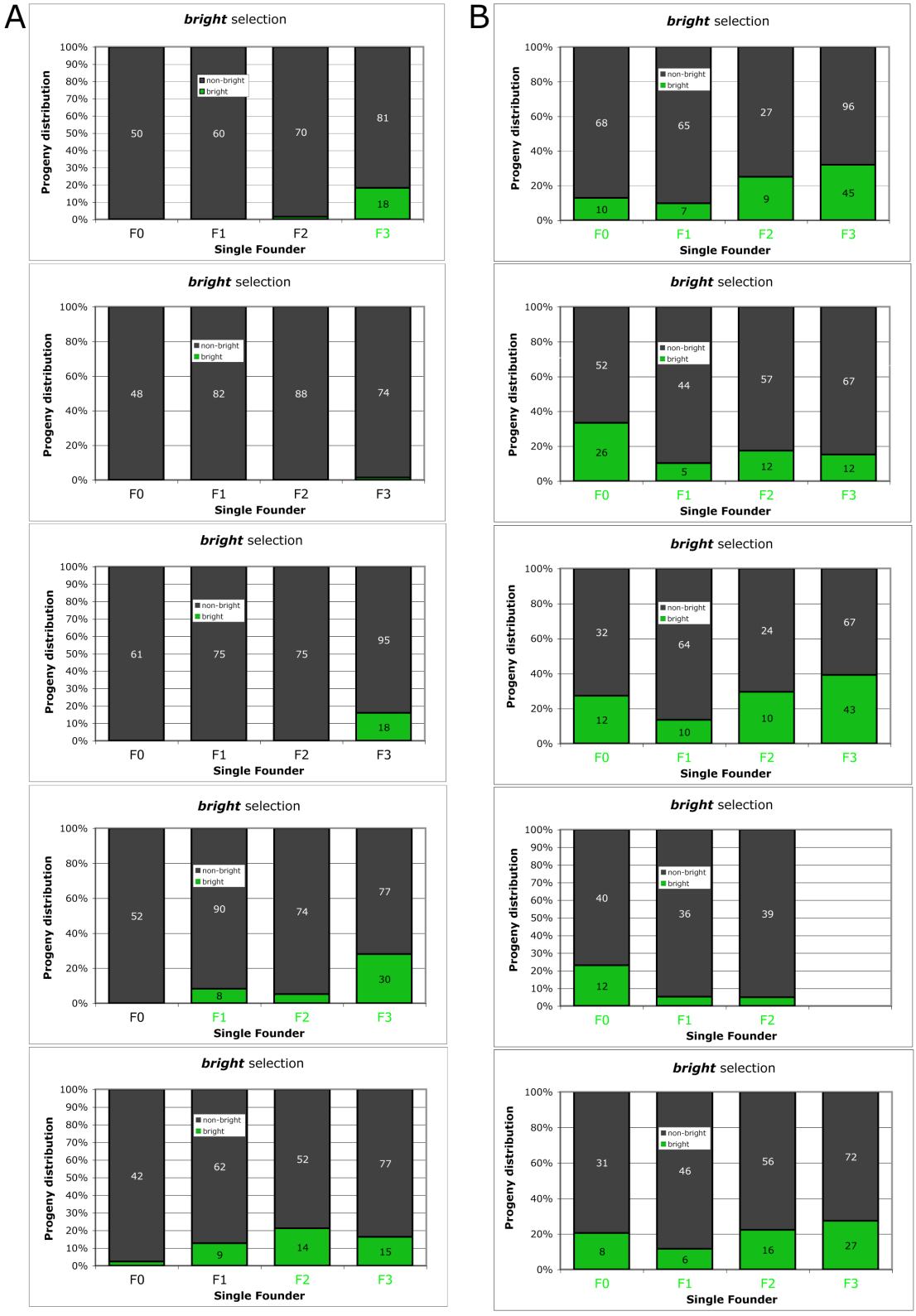


Figure 2.15. Selection for *bright*ness. (A) Each lineage was founded by a non-*bright* F0 animal. The color of the founder (black or green) indicates the phenotype of the founder to be non-*bright* or *bright*, respectively. In some cases, *bright* animals were not present in the population; so a non-*bright* founder was used instead. (B) Lineages derived from single *bright* F0 founders. In column B, there was always at least one *bright* animal present in any population to be used as the founder. Numbers in each column indicate the number of animals in that category.

Figure 2.15


A
B
bright selection
Progeny distribution
Single Founder
non-bright
bright
F0
F1
F2
F3
50
60
70
81
18
68
65
27
96
10
7
9
45
48
82
88
74
52
44
57
67
26
5
12
12
61
75
75
95
18
32
64
24
67
12
10
10
43
52
90
74
77
30
40
36
39
12
42
62
52
77
9
14
15
31
46
56
72
8
6
16
27

TABLES

Table 2.2. Summary of *bright* candidates from the second EMS screen. Multi-generation penetrance was determined by estimation of a non-starved, mixed population of animals. Single generation penetrance was determined by counting the fraction of *bright* progeny produced from a single hermaphrodite founder. Multiple determinations were made to produce the range of single generation penetrance shown below.

Table 2.2: *bright* candidates from 2nd EMS screen			
Candidate	**Penetrance**		**Comments**
	multi-generation	**single generation**	
PD3838	50-60%	33–96%	• dominant? • transgene not integrated
PD3841	50-60%	13–51%	• dominant?
PD3845	40-50%	29–80%	• dominant?
PD3844	60-90%	37–92%	• dominant? • *dpy unc*
PD3850	≥50%	0–79%	• dominant? • sick
PD3806	30-50%	25–60%	• dominant?
PD3901	30–50%	6–63%	• dominant?
PD3905	ND	0–95%	• dominant? • sick
PD3913	≈90%	8–53%	• confusing segregation pattern • phenotype not very strong
backburners	• 17 candidates with multi-generation penetrance less than 40% (most in 20-30% range)		

REFERENCES

BENDER, J., and G. R. FINK, 1995 Epigenetic control of an endogenous gene family is revealed by a novel blue fluorescent mutant of Arabidopsis. Cell **83:** 725-734.

BOWMAN, J. L., H. SAKAI, T. JACK, D. WEIGEL, U. MAYER *et al*., 1992 SUPERMAN, a regulator of floral homeotic genes in Arabidopsis. Development **114:** 599-615.

BRINK, R. A., 1958 Paramutation at the R locus in maize. Cold Spring Harb Symp Quant Biol **23:** 379-391.

CHANDLER, V. L., W. B. EGGLESTON and J. E. DORWEILER, 2000 Paramutation in maize. Plant Mol Biol **43:** 121-145.

COLLINS, J., B. SAARI and P. ANDERSON, 1987 Activation of a transposable element in the germ line but not the soma of Caenorhabditis elegans. Nature **328:** 726-728.

COLOT, V., L. MALOISEL and J. L. ROSSIGNOL, 1996 Interchromosomal transfer of epigenetic states in Ascobolus: transfer of DNA methylation is mechanistically related to homologous recombination. Cell **86:** 855-864.

DAS, O. P., and J. MESSING, 1994 Variegated phenotype and developmental methylation changes of a maize allele originating from epimutation. Genetics **136:** 1121-1141.

GRISHOK, A., H. TABARA and C. C. MELLO, 2000 Genetic requirements for inheritance of RNAi in C. elegans. Science **287:** 2494-2497.

JACOBSEN, S. E., and E. M. MEYEROWITZ, 1997 Hypermethylated SUPERMAN epigenetic alleles in arabidopsis. Science **277:** 1100-1103.

KETTING, R. F., T. H. HAVERKAMP, H. G. VAN LUENEN and R. H. PLASTERK, 1999 Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell **99:** 133-141.

MEYER, P., I. HEIDMANN and I. NIEDENHOF, 1993 Differences in DNA-methylation are associated with a paramutation phenomenon in transgenic petunia. Plant J **4:** 89-100.

MOERMAN, D. G., and R. H. WATERSTON, 1984 Spontaneous unstable unc-22 IV mutations in C. elegans var. Bergerac. Genetics **108:** 859-877.

MORGAN, H. D., H. G. SUTHERLAND, D. I. MARTIN and E. WHITELAW, 1999 Epigenetic inheritance at the agouti locus in the mouse. Nat Genet **23:** 314-318.

POTHOF, J., G. VAN HAAFTEN, K. THIJSSEN, R. S. KAMATH, A. G. FRASER *et al.*, 2003 Identification of genes that protect the C. elegans genome against mutations by genome-wide RNAi. Genes Dev **17:** 443-448.

RAKYAN, V. K., S. CHONG, M. E. CHAMP, P. C. CUTHBERT, H. D. MORGAN *et al.*, 2003 Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. Proc Natl Acad Sci U S A **100:** 2538-2543.

RASSOULZADEGAN, M., V. GRANDJEAN, P. GOUNON, S. VINCENT, I. GILLOT *et al.*, 2006 RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. Nature **441:** 469-474.

RICHARDS, E. J., 2006 Inherited epigenetic variation--revisiting soft inheritance. Nat Rev Genet **7:** 395-401.

STAM, M., C. BELELE, J. E. DORWEILER and V. L. CHANDLER, 2002 Differential chromatin structure within a tandem array 100 kb upstream of the maize b1 locus is associated with paramutation. Genes Dev **16:** 1906-1918.

VASTENHOUW, N. L., S. E. FISCHER, V. J. ROBERT, K. L. THIJSSEN, A. G. FRASER *et al.*, 2003 A genome-wide screen identifies 27 genes involved in transposon silencing in C. elegans. Curr Biol **13:** 1311-1316.

CHAPTER 3

Imprinting capacity of gamete lineages in *C. elegans*

This chapter appeared in whole in the following publication:

Sha and Fire (2005). Imprinting capacity of gamete lineages in *Caenorhabditis elegans*.

Genetics **170**: 1633-1652

INTRODUCTION

Parent-of-origin effects refer to a set of phenomena in which an entire set or subset of the paternal and maternal genome are distinguished from each other in the progeny genome. One of the first described cases of parent-of-origin effects was by Helen Crouse in 1960, who coined the term "imprinting" to describe the elimination of certain paternal chromosomes in *Sciara* flies (CROUSE 1960). Today, the term genomic imprinting is often used to describe the monoallelic expression of a gene from either the paternal or the maternal chromosome, but not from both. Genomic imprinting exists in a diverse set of organisms that span different phyla, including mammals, plants, insects, and fish.

Insects show diverse imprinting phenomena. Perhaps one of the more extreme forms of imprinting is found in the coccid insects, in which the entire paternal genome is epigenetically marked and silenced, rather than the silencing of individual paternal or maternal alleles. For example, the coccid mealybug does not possess sex chromosomes (BROWN 1959; BROWN 1961). Maleness in this group of insects is determined by the heterochromatization and elimination of the entire paternally-derived genome (BONGIORNI and PRANTERA 2003; BROWN and NELSEN-REES 1961). In the sciarid flies, the maternally-derived and paternally-derived chromosomes are distinguished from each other in the progeny genome. The somatic and germline development of these flies is driven by the selective loss of various paternal chromosomes in various tissues at different stages of the life cycle (GODAY and ESTEBAN 2001). In Drosophila, manipulation of chromosomal environments can sometimes result in previously non-imprinted genes now being expressed in a parent-of-origin manner (GOLIC *et al.* 1998;

LLOYD 2000). As in Drosophila, imprinting has not been shown to play a developmentally critical role in zebra fish; yet this organism has the capacity to methylate DNA in a parent-of-origin-specific pattern (MARTIN and MCGOWAN 1995).

Although flowering plant development is drastically different from animal development, genomic imprinting has been observed to be an important feature of the plant life cycle (ALLEMAN and DOCTOR 2000; SCOTT and SPIELMAN 2004; VINKENOOG *et al*. 2003). Reproduction in flowering plants is characterized by a unique double fertilization event. Each of two sperm nuclei, carried on the same pollen grain, fertilizes separate targets. One nucleus fertilizes the haploid oocyte to become the zygote; while the other sperm nucleus fertilizes the diploid central cell to become the endosperm, a source of nutrients for the embryo. Of the handful of genes found to exhibit a parent-of-origin effect in plants so far, all affect development of the endosperm. Two well-characterized imprinted genes in Arabidopsis are *MEDEA* (GROSSNIKLAUS *et al*. 1998) and *FWA* (KINOSHITA *et al*. 2004; SOPPE *et al*. 2000).

By far the most extensively studied examples of genomic imprinting have been in mammals. Early experiments involving translocations and nuclear transfer demonstrated the requirement for contribution of both parental genomes (i.e. CATTANACH and BEECHEY 1990; MANN and LOVELL-BADGE 1987). As of December 2004, over 70 murine genes have been listed on the Harwell Imprinting website to be imprinted (http://www.mgu.har.mrc.ac.uk/research/imprinting/imprin-viewdatagenes.html). Imprinting is critical for mammalian development, and defects in the imprinting process often lead to debilitating diseases (WALTER and PAULSEN 2003b). An interesting aspect of imprinting in the murine system is that many transgenes are also subject to imprinting.

From work by multiple labs over many years, certain themes have emerged concerning transgene imprinting in mammals. Generally, passage through the female germline results in decreased activity of the reporter transgene, ranging from partial (KEARNS *et al*. 2000; PREIS *et al*. 2003) to complete and irreversible silencing (i.e. LAU *et al*. 1999). Additionally, the expression imprint is correlated with DNA methylation levels, with the maternally-derived alleles generally being more methylated than the paternally-derived alleles.

Although reports of parent-of-origin effects in other organisms have been abundant, accounts of parent-of-origin phenomenon in *C. elegans* have been very rare. A screen for the requirement for biparental inheritance failed to uncover any evidence of whole-chromosome imprinting in *C. elegans* (HAACK and HODGKIN 1991). Kelly and colleagues recently reported the germline-sex-specific modification of the X chromosome in *C. elegans*. In their study, they observed a difference in chromatin state in the zygote between the spermatogenesis-derived and oogenesis-derived X chromosome. The differential chromatin marks persisted up to the 20-cell stage of embryogenesis (BEAN *et al*. 2004). In this article, we present evidence that a set of *unc-54* transgenes are expressed in a parent-of-origin manner in *C. elegans* and that the imprint persists into somatic development, but is reset upon passage through the opposite germline. Equivalent levels of expression are obtained when the transgene is transmitted through hermaphrodite sperm compared to transmission through male sperm, suggesting that whatever process(es) that act to establish and/or maintain the imprint is dependent upon the gamete type.

MATERIALS AND METHODS

C. elegans strains and growth conditions

Animals were reared on *E. coli* strain OP50 grown on NGM (nematode growth medium) nutrient plates according to standard protocols (BRENNER 1974). *C. elegans* strains used in the experiments were as follows:

N2: wildtype strain of *C. elegans* (Bristol isolate)

***pha-1*(*e2123ts*) III:** temperature-sensitive recessive mutation on Chromosome III (SCHNABEL and SCHNABEL 1990); homozygous *pha-1*(*e2123ts*) animals are viable at 16°C but embryonic lethal at temperatures above 20°C

***nIs106*[*lin-15*(+) + *lin-11*::*gfp*] X:** *lin-15::gfp* rescue line (REDDIEN et al. 2001)

***unc-119*:*gfp*(*edIs6*):** *unc -119::gfp* translation fusion (MADURO and PILGRIM 1995)

***ceh-23*::*gfp*(*lqIs27*):** (YANOWITZ et al. 2004)

tra-2*(*q122*) II:*** (SCHEDL and KIMBLE 1988)

PD3815 [*pha-1*(*e2123ts*) III; *ccEx3815*]: carries the extra-chromosomal tandem array *ccEx3815*[*unc-54*::*gfp* + *pha-1*(+)] in the *pha-1*(*e2123ts*) background

PD3852 and PD3872 [*ccIn3852 pha-1*(*e2123ts*) III]: integrated derivative of *ccEx3815*; *ccIn3852* is located at center of Chromosome III; PD3872 is PD3852 outcrossed once

PD3861 [*pha-1*(*e2123ts*) III; *ccIn3861* V]: integrated derivative of *ccEx3815*; *ccIn3861* is located at center of Chromosome V

PD3862 [*ccIn3862* I; *pha-1*(*e2123ts*) III]: integrated derivative of *ccEx3815*; *ccIn3862* is located at center of Chromosome I

PD3866 [*ccIn3862* I; *pha-1*(*e2123ts*) III; *him-5*(*e1467*) V]: PD3862 marked *in trans* with *him-5*(*e1467*)V

PD3920 [*dpy-5*(*e61*) *unc-13*(*e1091*) I; *pha-1*(*e2123ts*) III]: *dpy-5*(*e61*) *unc-13*(*e1091*) I; *pha-1*(*e2123ts*) III triple mutant

PD3924 [*ccIn3862 dpy-5*(*e61*) *unc-13*(*e1091*) I; *pha-1*(*e2123ts*) III]: PD3862 marked *in cis* with the two recessive mutations *dpy-5*(*e61*) and *unc-13*(*e1091*)

PD3928 [*unc-13*(*e1091*) I; *pha-1*(*e2123ts*) III]: *unc-13*(*e1091*) I; *pha-1*(*e2123ts*) III double mutant

PD3936 [*ccIn3852 dpy-17*(*e164*) *unc-69*(*e587*) *pha-1*(*e2123ts*) III]: PD3852 marked *in cis* with the two recessive mutations *dpy-17*(*e164*) *unc-69*(*e587*)

PD3938 [*ccIn3862 unc-13*(*e1091*) I; *pha-1*(*e2123ts*) III]: PD3862 marked *in cis* with *unc-13*(*e1091*)

PD3939 [*ccIn3862 dpy-5*(*e61*) I; *pha-1*(*e2123ts*) III]: PD3862 marked *in cis* with *dpy-5*(*e61*)

PD3940 [*ccIn3862 unc-13*(*e1091*) / *ccIn3862 dpy-5*(*e61*) I; *pha-1*(*e2123ts*) III; *him-5*(*e1467*) V]: *ccIn3862 unc-13*(*e1091*) / *ccIn3862 dpy-5*(*e61*) trans heterozygote in *pha-1*(*e2123ts*) and *him-5*(*e1467*) background

PD3942 [*ccIn3862 unc-13*(*e1091*) / *ccIn3862 dpy-5*(*e61*) I; *pha-1*(*e2123ts*) III]: *ccIn3862 unc-13*(*e1091*) / *ccIn3862 dpy-5*(*e61*) trans heterozygote in *pha-1*(*e2123ts*) background

PD3945 [*dpy-17*(*e164*) *unc-69*(*e587*) *pha-1*(*e2123ts*) III]: *dpy-17*(*e164*) *unc-69*(*e587*) *pha-1*(*e2123ts*) triple mutant

PD4251 [*ccIn4251* I; *dpy-20*(*e1282*) IV]: *ccIn4251*[*myo-3*::*gfp*(nuclear) + *dpy-20*(+)] (FIRE et al. 1998b)

PD8438 [*ccEx8438*; *pha-1*(*e2123ts*) III]: *ccEx8438*[*sur-5*::*gfp* + *pha-1*(+)] in *pha-1*(*e2123ts*) background

Map positions for integrations *edIs6* and *lqIs27* have not been determined. Strains were kept at either 16°C or 23°C, depending on whether or not they were *pha-1*(*e2123ts*) rescued. All genetic crosses were carried out at 23°C unless otherwise specifically indicated. In experiments for which the transgene array was linked to recessive genetic markers, each cross was closely monitored to ensure continual linkage between markers and the transgene array.

Plasmids and transgenic lines

A mixture of four plasmids was micro-injected (MELLO *et al.* 1991) into *pha-1*(*e2123ts*) animals. Plasmid pC1 (GRANATO *et al.* 1994) contains the wildtype genomic *pha-1* sequence without the 3' UTR. Plasmids pPD95.93, pPD105.21, and pPD120.90 each have a 204bp *unc-54* promoter segment driving the GFP coding region followed by the *unc-54* 3' UTR. These three plasmids differ in their combinations of subcellular localization signals: pPD95.93 carries a nuclear localization signal and *lacZ* (yielding nuclear GFP); pPD105.21 carries a mitochondrial localization signal; pPD120.90 carries four nuclear localization signals (yielding nucleolar GFP). Transgenic lines derived from this mixture show a relatively uniform pattern of GFP within expressing cells. In standard transgenic lines, the *unc-54* promoter provides mosaic

expression in body muscles of *C. elegans*. Several independent transgenic lines with these characteristics were obtained, and one, PD3815, was chosen for further analysis.

The transgene array in PD3815 is inherited as an extra-chromosomal element. Five confirmed integrated derivatives of this array were obtained following treatment with EMS (BRENNER 1974). Three of these integrations were mapped and were chosen for further analysis. These lines are designated PD3852, PD3861 and PD3862 and the corresponding integrated transgene loci as *ccIn3852*, *ccIn3861*, and *ccIn3862*.

Image capture

Animals of the desired genotype were sampled randomly and blindly (i.e. without knowledge of their GFP expression levels) and immobilized in mounting solution (50 mM NaCl, 5 mM EDTA, 0.5 mM levamisole) on eight-well glass slides (MP Biomedicals, Cat. #6040805). Fluorescent images of live animals were captured using a chilled CCD camera (Nikon CCD300ET-RC camera). Neutral density filters were used whenever necessary to ensure linearity of signal. All measurements were carried out in the linear range of detection as assayed by proportionality between observed signal and transmission percentage of the neutral density filter. Identical instrument and software settings were used in all image capture sessions.

GFP quantitation

Quantitation of GFP levels was done according to the procedure outlined in Figure 3.1. We used two sources of constant fluorescence in normalizing sample populations: a uniformly labeled fluorescent bead standard (Molecular Probes, Cat.#:M-7901) and a set of animals from a well characterized and stable GFP-expressing line (PD4251). Based on these values and those for the sample population, we obtain a

measure of GFP fluorescence. Analysis of fluctuations in the ratio between the reference line PD4251 and fluorescent bead standards provides a basis by which we can judge fidelity of the assay. In experiments carried out to date, the ratio between PD4251/beads is relatively constant (3.58±0.55; *n*=1,180), giving a variability of about 15% (standard deviation/mean).

Analysis of transgene DNA in transformed lines

DNA from strains PD3815, PD3861, PD3862, PD3872, and PD3924 was extracted, digested with restriction enzymes (*Pst I* + *Age I*, *Pst I* + *Nco I*, *Pst I* + *PspOM I*, *Mfe I* + *PspOM I*), separated by agarose gel electrophoresis, and visualized by ethidium bromide staining. Southern blot analysis was carried out using standard protocols. Briefly, electrophoresced DNA was blotted onto Hybond-N+ membranes (Amersham Biosciences, Cat. #RPN303B) according to the manufacturer's instructions. We used the method of capillary blotting under alkali conditions. Radiolabeled probes corresponding to a 222bp segment of the *unc-54* promoter were synthesized using the RadPrime DNA Labeling System (Invitrogen, Cat. #18428-11) with radiolabeled α-^{32}P dATP (MP Biomedicals, Cat. #:33002HD.5). Southern hybridization was done according to the protocol described in Molecular Cloning: A Laboratory Manual (RUSSELL and SAMBROOK 2001). We performed hybridization in roller bottles using phosphate-SDS buffer. Radioactivity corresponding to hybridized DNA fragments was transferred to a Molecular Dynamics phospho-screen followed by visualization and quantitation using a Molecular Dynamics phospho-imager.

Statistical analyses

The t-test analysis was performed using two-tailed distribution and two-sample unequal variances (PAGANO and GAUVREAU 2000). We additionally subjected the data sets in Figures 6 and 8 to computer simulations to determine if the observed trends could occur by pure chance alone (See Appendix).

RESULTS

Development of an assay for GFP quantitation of *C. elegans* populations

Much of the work in this paper involves comparing the relative GFP expression of genotypically identical populations of animals, the only difference being the parental source of the reporter chromosome (sperm versus oocyte) and the paired or unpaired character of the reporter-carrying chromosome in the parental generation. Measurement of relative GFP levels between two populations of animals presents a challenge. We obtained quantitative fluorescence data first by acquiring digitized images of each animal using a CCD camera system. Based on the experimental measurement of total fluorescence integrated over each sample image, we obtain a total signal. These signals are then corrected, as appropriate, by background subtraction to obtain an essentially quantitative comparison of different animal populations (Figure 3.1). Although some variation is inherent in this assay, we consider the results of the assay to be accurate to within an approximate variance of 15%.

A fusion reporter for quantitative expression analysis

We initially set out to screen for parent-of-origin effects on gene expression in *C. elegans* using an *unc-54*::*gfp* transcriptional fusion reporter assay. The *unc-54* gene in *C. elegans* encodes the major myosin heavy chain of bodywall muscles (BRENNER 1974; EPSTEIN *et al.* 1974; MACKENZIE *et al.* 1978). Briefly, plasmids containing the *unc-54* promoter fused to GFP were micro-injected with a genomic clone of the *pha-1*(+) gene into L4 animals homozygous for the *pha-1(e2123ts)* mutation (GRANATO *et al.* 1994). At 16°C, the *pha-1(e2123ts)* mutation is permissive; at 23°C, 100% of *pha-1(e2123ts)* animals arrest as embryos or L1 larvae.

Plasmid DNA populations injected into *C. elegans* are subject to a recombination process that in many cases leads to the formation of long, extra-chromosomal tandem arrays that can be inherited from generation to generation in a non-Mendelian manner (STINCHCOMB *et al.* 1985). When two or more plasmids are mixed, the extra-chromosomal transgene array consists of tandem mixed arrays of the two sequences, with a total copy number of 100-200. By rearing populations at 23°C, we select for *pha-1*(+) animals that harbor the transgene array. A single clonal line from the plasmid injections was selected for the initial analysis. This line, designated PD3815 (Figure 3.2) is viable at 23°C and exhibits detectable body wall muscle GFP activity.

For reasons that are not well understood, a large fraction of transgene arrays in *C. elegans* exhibit mosaic expression: expression in some but not all of the cells where expression would be expected based on knowledge of the promoter used in the initial reporter construct (FIRE and MELLO 1995). Although a fraction of the mosaicism is due to mitotic loss of the extra-chromosomal array, substantial mosaicism is observed even in the majority of transgenic lines harboring arrays that have integrated into the chromosome and thus should be present in every cell. Patterns of expressing and non-expressing cells tend to be random within a target tissue, with little or no adherence to lineal boundaries. The seemingly random pattern of GFP activity from array-based transgenes further argues against the mitotic loss of arrays as the sole cause of mosaicism. Instead, it appears that arrays are physically present but transcriptionally silenced in a large number of cells. Consistent with the hypothesized gene silencing, the level of reporter expression in transgenic strains carrying multi-copy arrays is rarely a simple multiple of expression observed with rare single copy integration events. Rather it

appears that the arrays are essentially in a silent state with rare copies (or whole arrays) undergoing a rare activation event.

To obtain a strain with a greater degree of uniformity than the original transgenic PD3815, we carried out an EMS mutagenesis, followed by a screen one to three generations later for animals with more uniform and increased level of GFP expression. Among a collection of candidates, we pursued the three with strongest expression. The resulting lines, designated PD3852, PD3861, and PD3862, exhibited increased overall expression combined with somewhat reduced mosaicism (Figure 3.2). Conceivably, the improved GFP activity in these three strains could result from a variety of different alterations. In the case of PD3852, PD3861, and PD3862 the critical alteration appears to be in the structure or genomic context of the transgene locus and not the induction of mutations in the genetic background that facilitate transgene expression. In particular, the array in each of these lines had become integrated into the genome. Integration of arrays is a very unusual event in *C. elegans* (FIRE *et al.* 1991; MELLO *et al.* 1991) and three independent unselected integrations would be vanishingly rare. It should be noted that not all integration events lead to the type of expression improvements seen with PD3852, PD3861, and PD3862 (HSIEH *et al.* 1999). The integration loci in lines PD3852, PD3861, and PD3862, designated *ccIn3852*, *ccIn3861*, and *ccIn3862*, respectively, map to the central regions of three distinct chromosomes (*ccIn3852*: chromosome III, *ccIn3861*: chromosome V, *ccIn3862*: chromosome I). Conceivably, these could be specific genomic regions that allow localized transgene de-silencing; alternatively, the precise structure of the three integrated arrays may have been selected to promote the observed relief of silencing. In either case, the localized nature of the de-silencing response is supported by

genetic mapping, from which uniform GFP expression in each line maps solely to the integrated array. No evidence for second site activating mutations was obtained in outcross experiments.

Southern blot analyses similar to those of Stinchcomb et al (1985) were used to examine structures of the original extra-chromosomal transgene as well as the three integrated derivatives (see methods). As expected each of these lines contains a tandem array with sequences from *unc-54*::*gfp* and the *pha-1*(+) marker. Each array appears to contain 20-30 *unc-54*::*gfp* copies interspersed among 50-200 copies of the *pha-1*(+) marker plasmid. The *pha-1*(+) marker was clearly visible in the genomic digests of *ccEx3815*, *ccIn3852*, *ccIn3861*, and *ccIn3862* (Figure 3.3A), indicating high copy numbers. Although no differences in copy number or array structure were evident following restriction digestion with a representative group of enzymes, we cannot rule out the possibility that point mutations or subtle changes in configuration might have an impact on expression for the different integrated loci. Similar analysis of insertions *edIs6*(PD2169) (MADURO and PILGRIM 1995) and *ccIs9385*(PD9385) (FIRE *et al.* 1998a) demonstrated that each of these lines also carried a high-copy-number tandem array.

Unexpected expression ratios from simple outcrosses

While out-crossing the three integrated *unc-54*::*gfp* strains from the mutagenized genetic backgrounds, we found an anecdotal correlation in some cases between parent-of-origin for the transgene and expression level. We further characterized this effect in outcrossed lines in experiments where a single transgene locus was introduced from either oocyte or sperm into a zygote, followed by quantitation of expression levels in the larval and adult stages. As shown in Figure 3.4A, we observed that two of the

transgene loci (*ccIn3852* and *ccIn3862*) expressed more strongly when the transgene locus was introduced from the male parent (i.e. through sperm). Expression was still observed following transmission through oocytes, but was seen at a significantly lower level. No evidence for a parent-of-origin effect was observed with the third array, *ccIn3861*. The parent-of-origin effect was a stable property of lines PD3852 and PD3862: consistent results were obtained with these lines in experiments carried out in several different generations following the establishment of the outcrossed line, while no parent-of-origin-dependent expression was observed at any point with PD3861 (Figure 3.4A). In addition to the observed parent-of-origin effect, we also observed an unexpectedly skewed ratio when comparing expression between homozygotes and oocyte-derived hemizygotes (Figure 3.4B). These skewed ratios presumably reflect the parent-of-origin effect combined with other (yet-to-be-characterized) consequences of homozygosity for the transgene.

Both integrated and extra-chromosomal transgenes exhibit parent-of-origin expression patterns. In particular, the ratio of sperm-derived to oocyte-derived expression of *ccEx3815* in line PD3815 was 2.2 (data not shown). Thus, integration is not a pre-requisite for the imprinting of the *unc-54*::*gfp* transgene, since the extra-chromosomal array *ccEx3815* is also imprinted.

Several different processes could skew the ratio of expression levels in these experiments. There is certainly precedent for chromosomal imprinting from other systems. This would entail a mechanism in which the maternal and paternal chromosome sets provide different expression levels to the zygote. As an alternative, however, one must consider that the cytoplasmic contributions in the two crosses shown in Figure 3.4

are distinct. In the case of maternal acquisition of the transgene, the maternal cytoplasm is derived from an animal carrying the transgene. Products of the transgene locus (such as modulatory or aberrant RNAs) might be present in the oocyte and thereby modulate the subsequent activity of the reporter construct. Fortuitous transcripts of transgene loci have frequently been observed (FIRE *et al*. 1991) and could potentially produce stable changes in gene expression, so this possibility cannot be ruled out without clear experimental tests. A second potential difference between the two crosses in Figure 3.4 relates to the meiotic pairing state of the transgene array in the parents of the assayed animals. In one case the parent is homozygous for the transgene array, hence an opportunity (at least) for pairing of the locus in the previous generation. In the other case the parent is hemizygous, suggesting the possibility of an unpaired state. Meiotic silencing of unpaired DNA has been demonstrated in Neurospora (ARAMAYO and METZENBERG 1996) and has been suggested for *C. elegans* (BEAN *et al*. 2004).

Expression of *ccIn3862* depends on the gamete-of-origin for the transgene chromosome

The suggestion of parent-of-origin effects with *ccIn3852* and *ccIn3862* led us to further investigate the genetic basis in greater detail. To determine the source of the observed skewing of expression ratios, we set up a series of crosses in which four variables were tested: parent of origin of the transgene array, pairing of the array in the parental generation, cytotype (cytoplasmic genotype of parental germline), and hemizygosity/homozygosity of the array in assayed progeny. We organized these experiments into the matrix shown in Figure 3.5. Eight different crosses yielded animals that were hemizygous for the transgene array while four crosses yielded array

homozygotes. Each intersection is an independent experiment; hence 12 independent experiments were carried out to complete the matrix analysis. Note that we were careful to keep the genotypes of the hermaphrodite parents (rows) and male parents (columns) as consistent as possible (Figure 3.5B). We chose to analyze *ccIn3862* over *ccIn3852* because of the relative ease of constructing the necessary genetically marked derivatives of *ccIn3862*. *ccIn3861* was not included in the analysis because it does not show a parent-of-origin effect.

The central question of our matrix experiments was whether, on average, two genotypically identical populations, the only difference being the parent-of-origin or pairing state in the parental germline of *ccIn3862*, differ in their level of GFP expression. The results of the matrix experiments are shown in Figure 3.6, with the accompanying statistical analysis. Each value in the matrix is a population average from numerous animals. Comparing each diametrically opposite pair (for example A3 versus C1), we saw that in all cases transmission from sperm gave greater expression than transmission from oocyte. This difference is statistically significant in all cases (Figure 3.6C) and for both adult and L4 animals alike. Significantly, the matrix data recapitulate results from a set of earlier experiments (Figure 3.4B) that detect non-linear expression ratios between array homozygotes and hemizygotes. In the adult data set in Figure 3.6, GFP expression of array homozygotes (C3, C4, D3, D4) are at least twice that of any single array derived from hemizygous parents (C1, C2, A3, B3), although this difference is slightly smaller when the single array was derived from homozygous parents (A4, B4, D1, D2). The L4 data set show a similar trend.

We generated scatter plots of normalized expression values (i.e. $x_{e,i\ (normalized)}$) to determine the spread of the data points (Figure 3.7). As can be seen from the plots, each population displayed a range of GFP expression. Although there were outliers, the majority of the data points clustered around the population mean. It appears that for any given data set, L4 animals show less scattering than adults. This is probably due to the L4 being a more defined stage of development than the adult stage.

ccIn3862 is expressed equivalently from male and hermaphrodite sperm

Because *C. elegans* sperm can be derived from either males or hermaphrodites, an intriguing question is whether expression of *ccIn3862* when derived from male sperm is equivalent to its expression when derived from hermaphrodite sperm. We sought to answer this question genetically by performing the experiment shown in Figure 3.8. Animals from progeny class A received *ccIn3862* from the oocyte; whereas class B animals received *ccIn3862* from male sperm. Class C animals were derived from selfing of F1 parents hemizygous for *ccIn3862*. Statistically, half of class C animals should receive *ccIn3862* from the oocyte and half should receive *ccIn3862* from hermaphrodite sperm. We reasoned that since *ccIn3862* in class C is a mixture of both sperm-derived and oocyte-derived, if male and hermaphrodite sperm expressed *ccIn3862* equivalently, then the population average of class C should be intermediate, larger than that of class A but smaller than that of class B. This prediction is indeed borne out (Figure 3.8D). Additionally, when we combined classes A and B into a single class and determined the population average of the combined AB class, the combined AB average is indeed statistically equivalent to the average of class C. From these observations, we conclude that with respect to expression of *ccIn3862*, we do not see evidence for a difference

between male versus hermaphrodite sperm transmission of *ccIn3862* (Figure 3.8D and E).

Lack of an observed pairing effect on *ccIn3862* expression

It has been proposed in several systems that the pairing state of a locus in parental meiosis can be a determinant in setting the expression level of the subsequent generation. The experiments described in Figure 3.5 provide an initial indication of the relative contributions of pairing history and parental origin to expression level for the *ccIn3862* transgene. Although the strongest apparent effect on expression of this transgene is from a parent-of-origin effect, an additional effect of pairing history needs to be considered. The data provide evidence that if such a pairing effect influences the expression of *ccIn3862*, this effect is rather limited in magnitude. The slight expression advantage among genotypically identical populations where *ccIn3862* was transmitted from a homozygous state is within the margin of "noise" that we generally observe in these assays (Figure 3.6A and B, columns C versus D and rows 3 versus 4). To obtain an additional measure of potential effects of pairing history on *ccIn3862* expression, we performed the experiment shown in Figure 3.9. In Figures 3.9A-C, we compared genotypically identical populations. The only difference between these three populations was the context from which *ccIn3862* was transmitted from parent to progeny. In Figure 3.9A, the male parent carried two *unc-54*::*gfp* transgenes integrated on different chromosomes: *ccIn3862* on chromosome I and *ccIn3852* on chromosome III. However, only progeny carrying *ccIn3862* were used for fluorescence measurements. In Figure 3.9B, the male parent was homozygous for *ccIn3862*. In Figure 3.9C, the male parent was hemizygous for *ccIn3862*. Hence, the situation we have set up was a comparison of

unc-54::*gfp* expression where the status of the array in the parental generation was two loci unpaired (Figure 3.9A) versus one locus paired (Figure 3.9B) versus one locus unpaired (Figure 3.9C). The results showed at most marginal differences between animals derived from the three crosses (within the 15% "noise" window which we generally accord the assays). These observations (and those of Figure 3.6) do not rule out pairing effects on transgene expression in *C. elegans*; rather they indicate that any potential pairing effect on this particularly late-expressed transgene are relatively modest in magnitude.

Resetting of the *ccIn3862* imprint after long term maintenance in a single gamete lineage

The experiments described up to this point have shown that resetting of the transgene can occur after a single-generation passage through the opposite germline. In these experiments, the transgene array had been initially kept in the hermaphrodite parent: it is only at the imprinting cross (where we performed reciprocal crosses) that we distinguished between the oocyte versus sperm source of the transgene array. Since *C. elegans* hermaphrodites produce both sperm and oocytes, we would expect that a transgene kept in hermaphrodites has an equal chance at each generation of being sperm-derived or oocyte-derived. We thus expect the state of the hermaphrodite-derived transgene in such populations to reflect (in the long run) an average of the oocyte and sperm-derived values. This is indeed seen in the experiment described in Figure 3.8, where the hermaphrodite-derived value (3.90±1.67, cross C) is equal to the average of the oocyte plus sperm-derived values (3.76±1.96, crosses A plus B).

It is possible using appropriate genetic markers to engineer the continued passage of a locus through a single germ line (oocyte or sperm) for many generations. Several long-term genetic experiments to test the effects of long term oocyte or sperm transmission were carried out using transgene *ccIn3862* (Figure 3.10). Figures 10A and 10D show one example of expression levels from such an experiment. For the first 44 generations *ccIn3862* was transmitted through the oocyte or sperm lineage for 10-12 consecutive generations in each germline. For the next eight generations (F45-F52), transmission was alternated at each generation (oocyte-sperm-oocyte-sperm). Throughout the entire experiment, the array was transmitted as a hemizygote (Figure 3.10B). All animals assayed had genotype *ccIn3862 dpy-5*(*e61*) *unc-13*(*e1091*)/+; *pha-1*(*e2123ts*). The only difference between assayed populations was the source of the transgene array (i.e. oocyte-derived versus sperm-derived) and number of generations through each germline (i.e. one versus ten generations through oocyte or sperm).

The results in Figure 3.10 illustrate several properties of the gamete-specific effect. Most strikingly, the gamete-specific effect is somewhat cumulative. This was particularly striking for sperm transmission, for which expression increased for multiple generations during continued passage (Figure 3.10D, F13-F23; F34-F39). In later generations, increases in expression appeared to slow or plateau (Figure 3.10C F29-F45; Figure 3.10D F39-F44). A multi-generational effect through oocyte transmission was also indicated, although this effect may require fewer generations than the maximal sperm effect (Fig 9C F46-F51; Figure 3.10D F24-F33).

All of these experiments indicate that *ccIn3862* that has been continually transmitted in one gamete line can acquire a state with some degree of meiotic stability.

The locus thus appears to retain some memory of meiotic source reaching back at least a handful of generations. This has some interesting quantitative consequences. In particular, the results of long term passage through the same germline followed by a single "switched" generation can result in average values that are different from those starting from mixed populations of sperm- and oocyte-derived transgene loci (i.e. those in Figure 3.6). For example, following long term sperm transmission and a single generation of oocyte transmission (Figure 3.10C: F46, Figure 3.10C: F24 and F45), the locus could retain activity higher than for some of the earlier sperm-derived values (i.e. Figure 3.10D: F34-F36).

Some, but not all, additional transgenes are imprinted

We tested several additional GFP transgenes to see if they were also subject to a parent-of-origin effect. Of five non-*unc-54* transgenes tested, only *unc-119*:*gfp*(*edIs6*) exhibited a parent-of-origin effect (Figure 3.11). *unc-119*:*gfp*(*edIs6*) is a translational fusion that is seen strongly in the nervous system and faintly in a number of additional tissues (MADURO and PILGRIM 1995). This construct shows a reproducible difference when comparing animals with a sperm-derived transgene locus (higher expression) with animals that carry an oocyte-derived locus (lower expression).

***In silico* validation of experimental data**

We used computer simulation to test the observed statistical significance of data presented in Figure 3.6 (matrix experiment) and Figure 3.8 ("3862SE", bottom four rows). Figure 3.12 illustrates the methodology. Two experimental data sets whose averages were to be compared were pooled. The pooled data set was divided into two random data sets; each randomized data set contained the same number of data points as

the original two data sets. Comparison of the averages of the two randomized data sets allowed us to determine statistical significance between the two original sets. If the difference between the two original data sets was statistically significant, then the averages of the randomized data sets should lie between the averages of the two original data sets (i.e. the averages of the randomized data sets should not equal the averages of the experimental data sets). Each paired data set was subjected to 10,000 iterations. Each iteration consists of the pooling of the two experimental data sets, generation of the two randomized data sets from the pool, and determination of the averages of the two randomized data sets. The computer simulation results show the observed differences to be statistically significant for all the paired data sets listed, except for the bottom row, where AB is supposed to be statistically equivalent to C (Figure 3.8).

DISCUSSION

In this study, we describe our analysis of parent-of-origin-specific imprinting with a set of GFP reporter transgenes in *C. elegans*. For one well studied set of transgene constructs (*unc-54*::*gfp*) we observed imprinting in two of three independently integrated transgenic strains as well as in the progenitor extra-chromosomal array from which the integrated lines were derived. Since imprinting effects on transgene expression in *C. elegans* had not been reported, these results were somewhat surprising. Further analysis of additional transgenic loci (integrated lines bearing different reporter constructs) indicated that the ability to imprint, although not widespread, was not limited to a single transgene construct or integration site.

The genetic analysis of imprinting for *unc-54*::*gfp* transgenes clearly indicates that *C. elegans* oocyte and sperm lineages have the capacity to differentially imprint parental chromosomes. Unlike certain imprinting events in other systems in which one parental allele is completely silenced, the imprinting we observed in our study was not an "on/off" situation. Because we are characterizing incomplete imprinting of a reporter transgene, our analysis is necessarily quantitative in nature. For an exemplary transgene, *ccIn3862*, we found 1.5-2.0 fold greater expression in progeny that received the transgene from their fathers (sperm transmission) as compared to those receiving the same transgene from mothers (oocyte transmission). The most definitive imprinting assays are those in which activities of a single genetic locus (in this case a transgene) are compared under conditions where maternal and paternal "cytoplasmic" contributions are kept constant (each is hemizygous for the relevant transgene). We were able to take advantage of *C. elegans* genetics to construct such a situation. Thus, we could dissect a

parent-of-origin effect under conditions where only the sperm or oocyte source of the transgene was varied.

In mammals, about 80% of imprinted endogenous genes occur within clusters with other imprinted genes (REIK and WALTER 2001; VERONA *et al*. 2003). The occurrence of imprinted genes close together has been proposed to reflect coordinate control of these genes by a central imprint control region (ICR). For many imprinted genes in mammals, differential DNA methylation is observed at a CpG-rich region called the DMR (differentially methylated region). Studies in mice indicate a requirement for the DMR and other sequences in the proper temporal and spatial control of imprinted gene expression (i.e. AINSCOUGH *et al*. 1997; REINHART *et al*. 2002; THORVALDSEN *et al*. 1998; WUTZ *et al*. 1997). Although the idea of coordinate control by an imprint control region seems elegant, the real picture is not so clear. First, what exactly marks a sequence of DNA for imprinting is not known. In the majority of imprinted clusters, certain genes within the cluster escape imprinting, and other genes within the cluster are imprinted only in specific tissues or at specific developmental stages. These observations suggest the existence of *cis*-acting sequences which may protect the genes from an imprinting effect or direct the temporal or spatial expression of an imprinted gene. Additional studies using transgenic mice indicate that integration into an imprinted cluster is not a requirement for the transgene to exhibit a parent-of-origin effect (KEARNS *et al*. 2000). Given these complexities, one expects that the determination of an imprinted status of a given DNA sequence is likely a combination of many mechanisms. Studies in multiple genetic organisms have clearly demonstrated that local structure (either covalent or epigenetic) plays a critical role in genetic state. This is particularly readily observed with transgene

insertions, where (at least for plants and mice) different copy numbers and/or array structures for the same transgene locus can display different levels of expression (i.e. DAY *et al*. 2000; GARRICK *et al*. 1998).

For unknown reasons, certain transgene constructs (including those driven by a minimal *unc-54* promoter) are particularly susceptible to silencing in *C. elegans*. Thus, we were surprised to obtain three integrated derivatives of an *unc-54*::*gfp* transgene (*ccIn3852*, *ccIn3861*, *ccIn3862*) in which silencing had apparently been partially or completely lifted. Two of these (*ccIn3852* and *ccIn3862*), along with the progenitor *ccEx3815* construct, exhibit an imprinting effect (Figures 3.4 and 3.5). The observed lack of a parent-of-origin effect for *ccIn3861* could conceivably reflect either the internal structure of the transgene array (*cis*-acting sequences which render *ccIn3861* resistant to imprinting), position in the genome (i.e. the integration site of *ccIn3861* may not be susceptible to imprinting), or an insufficiently sensitive assay. In light of the fact that the extra-chromosomal *ccEx3815* is also imprinted, it seems less likely that a *cis*-acting sequence required to confer imprinting is present in the *unc-54*::*gfp* constructs that exhibit a parent-of-origin effect; instead, it may be that *ccIn3861* has acquired a resistance to imprinting by virtue of its chromosomal environment. As all of our results are based on quantitation, it is certainly conceivable that a subtle parent-of-origin effect may have existed in *ccIn3861* but was below the detection threshold of our assay.

Work in Drosophila has demonstrated that modulation of the chromosomal environment of certain genes can lead to these genes acquiring a parent-of-origin effect. Numerous inversions, translocations, and duplications have taken groups of genes out of their endogenous context and inserted them nearby or into heterochromatic regions. The

result being that, in some cases, all genes on the displaced chromosomal segment now acquire a parent-of-origin effect (LLOYD 2000). Work by Lloyd and others have shown that, among a group of displaced genes, the closer a displaced gene is to heterochromatin, the greater is its degree of imprinting compared to other more distal genes in the same group (COHEN 1962; LLOYD *et al*. 1999). Maggert and Golic have shown that the entire Y chromosome of Drosophila can confer an imprinting status to transgenes (MAGGERT and GOLIC 2002). Hence, in Drosophila, observed imprinting is invariably associated with heterochromatin.

Parent-of-origin imprinting may not be the only mechanism that modulates expression of transgene loci. In particular, we have found that homozygote expression of *ccIn3852*, *ccIn3861*, and *ccIn3862* is somewhat greater than the expected sum of expression following independent sperm plus oocyte transmission (Figure 3.4B). Perhaps the non-linear expression ratio is due to a combination of parent-of-origin and pairing effects. DNA-DNA pairing is an important feature of gene silencing in other systems, as seen with RIP and MSUD in Neurospora (ARAMAYO and METZENBERG 1996; CAMBARERI *et al*. 1989), transvection and pairing-dependent silencing in Drosophila (i.e. DORER and HENIKOFF 1997; HENIKOFF and DREESEN 1989; LEWIS 1954) and plants (ASSAAD *et al*. 1993; MATZKE *et al*. 1994). Work by Bean and colleagues (2004) suggest pairing may be a critical feature of chromatin marks in the early *C. elegans* embryo. Although we have not observed a strong pairing effect on the later-expressed *unc-54*::*gfp* transgene described herein, it is important to note that a modest or earlier effect could have been missed.

In order to achieve monoallelic expression of a gene, the parents must establish imprints that mark the two parental alleles as distinct, and the progeny must then maintain the imprints in the somatic cell lineages (DELAVAL and FEIL 2004). Work by Tucker *et al.* indicate that (at least in mammals) germline passage is a requirement for the establishment and proper expression of imprinted genes (TUCKER *et al.* 1996), suggesting that establishment of the imprint occurs during gametogenesis where oogenesis and spermatogenesis differentially mark the maternal and paternal alleles, respectively. In mammals, this is evidenced by differential methylation at the DMRs. Non-histone DNA binding proteins as well as *cis*-acting sequences are important for directing differential methylation during oogenesis and gametogenesis (FEDORIW *et al.* 2004; PANT *et al.* 2003; PERK *et al.* 2002; SCHOENHERR *et al.* 2003; YOON *et al.* 2002). Differential chromatin modifications also play an important role in the establishment of imprints (i.e. XIN *et al.* 2001). Likewise, maintenance of the imprint (at least in mammals) likely involves multiple mechanisms, including maintenance of methylation by *Dnmt1* (BESTOR 2000; HOWELL *et al.* 2001), protection of unmethylated sites by DNA-binding proteins (FEDORIW *et al.* 2004; PANT *et al.* 2003; SCHOENHERR *et al.* 2003), differential chromatin modifications (FOURNIER *et al.* 2002; GREGORY *et al.* 2001; YANG *et al.* 2003), *Polycomb* group proteins (MAGER *et al.* 2003; OTTE and KWAKS 2003), and potentially non-coding RNAs (FITZPATRICK *et al.* 2002; SLEUTELS *et al.* 2002).

While DNA methylation is an important feature of epigenetic gene silencing in mammals, plants, and fungi, DNA methylation has not been found to exist in *C. elegans*. The lack of DNA methylation does not preclude *C. elegans* from gene silencing activities. Genetic data in Drosophila are consistent with the hypothesis that histone

modification serves the necessary role of localized information storage during imprinting (JOANIS and LLOYD 2002). By analogy, our current working model is that the imprinting of *C. elegans* transgenes likely involves the establishment of metastable histone modification states during gametogenesis and the subsequent maintenance and expression of these epigenetic states during embryonic proliferation. Differences in chromatin state would conceivably result from a combination of activating histone modifications upon passage of the transgenes through the sperm and/or de-activating modifications upon transmission through the oocyte.

Because *C. elegans* is a hermaphroditic species, hermaphrodites undergo both spermatogenesis and oogenesis; whereas males undergo only spermatogenesis. Germline development in both sexes occurs under a program of temporal and spatial separation (L'HERNAULT 1997; SCHEDL 1997; SINGSON 2001). The cytological processes of spermatogenesis and oogenesis are quite distinct. Spermatogenesis in both males and hermaphrodites occurs as a meiotic precursor cell in the gonad undergoes two rapid divisions, leaving the bulk of cytoplasm behind to produce four very compact spermatids. Hermaphrodite and male sperm are different, with male sperm larger in volume by two-fold (LAMUNYON and WARD 1998). Oocytes in *C. elegans* are large (50 μm x 30 μm x 30 μm) cells that have a 4n DNA content. Meiosis is completed in this species only after fertilization, with the meiotic spindle serving as the organizing center in generating cellular polarity. Genetic screens have identified diverse molecular players in gameto-genesis and meiosis, some of which are sperm-specific, some oocyte specific, and some involved in both processes (i.e. HODGKIN *et al.* 1979; REINKE *et al.* 2000). In general, oogenesis and spermatogenesis seem to have more unique than shared features. Most

components that play a role in spermatogenesis are required in both hermaphrodites and males, although there are a small number of exceptions required only in hermaphrodites (L'HERNAULT *et al.* 1988).

Our data indicate that expression of *ccIn3862* when it is transmitted through male sperm is equivalent to its expression when it is transmitted through hermaphrodite sperm (Figure 3.8). This result suggests that it is the sex of the gamete (i.e. oocyte versus sperm) and not the chromosomal or physiological sex of the parent that is critical in establishing/maintaining an imprinted state of the transgene. Although hermaphrodite sperm and male sperm differ in size and competence for fertilization (male sperm out-compete hermaphrodite sperm for fertilization (LAMUNYON and WARD 1997)), our result suggests that the process of establishment and/or maintenance of this imprint is not substantially different between hermaphrodite sperm and male sperm. Interestingly, Bean *et al.* (2004) also found that the imprinted chromatin state of the X chromosome in early embryos of *C. elegans* is also dependent upon the sex of the gamete.

Maintenance of a constant expression profile in a population exhibiting genetic imprinting will be most effective if the organism has a mechanism to reverse any imprint in the subsequent generation. Numerous examples, particularly from studies in mammals, have demonstrated that passage of an imprinted transgene locus through the opposite germline for a single generation results in the resetting of the imprint (REIK *et al.* 1987; SAPIENZA *et al.* 1987; SWAIN *et al.* 1987). Although the relief of imprinting in subsequent generations may be the rule, there are exceptions where the imprint appears to be meiotically stable (i.e. HADCHOUEL *et al.* 1987; KEARNS *et al.* 2000; LAU *et al.* 1999; RAKYAN *et al.* 2003). In mice, maternal transmission of an imprinted locus is generally

associated with decreased gene activity and hypermethylation of the transmitted DNA sequence. Comparison of the methylation status of DNA from somatic and germline tissues of male mice who had inherited their transgenes from their mothers indicated that DNA in somatic tissue was more heavily methylated than DNA in sperm (REIK *et al.* 1987; SAPIENZA *et al.* 1987; SWAIN *et al.* 1987). Although the methylation status of the oocyte was not determined, these studies indicated that the methylation status of the maternal parent was transmitted to the somatic cells of the male progeny, but was erased during male gametogenesis. Martin and McGowan recapitulated these findings in their work with transgenic zebra fish (MARTIN and MCGOWAN 1995). Chaillet and colleagues provided evidence that a murine transgene had lost its parent-specific methylation patterns in primordial germ cells in both sexes. Sex-specific patterns began to emerge during both oogenesis and spermatogenesis. In the female germline, a female-specific methylation pattern was completely re-acquired by late oogenesis; whereas male-specific patterns were completed only after fertilization (CHAILLET *et al.* 1991).

Similar to the murine and zebra fish examples, we have observed repeatedly that the imprinted status of *C. elegans* transgenes can be at least partially reset in a single generation after passage through the opposite germline. Hence, as in the mammalian examples, gametogenesis in the opposite germline can be sufficient to reset the imprint. *C. elegans* apparently has a concerted mechanism for reactivation of transgenes during somatic development (HSIEH *et al.* 1999). This mechanism is most evident in examining a set of mutant strains (i.e. *tam-1* loss of function) in which expression of tandem array transgenes fails to reactivate in somatic lineages. It is certainly conceivable that the resetting of the parent-of-origin imprint during gametogenesis could share components or

mechanistic features with the subsequent reactivation of transgene expression in somatic lineages.

Earlier observations with germline-expressed transgenes in *C. elegans* demonstrated that many are subject to a progressive and meiotically stable gene silencing process over the course of several generations (KELLY and FIRE 1998). The experiment shown in Figure 3.10 demonstrates that a meiotically stable state can be established for either a deactivated or an activated state through long term transmission through the oocyte or sperm lineage, respectively. Furthermore, the meiotically stable state can be reversed by multi-generation passage through the opposite germline; the extent of reversal being a function of the amount of time the transgene array experiences gametogenesis in the opposite germline. This implies that gametogenesis in each generation is a fixed window of time in which each germline establishes its unique epigenetic marks. A very stable epigenetic state, therefore, would require multiple generations to establish or reverse. In *C. elegans*, the *mes* genes (KELLY and FIRE 1998) and the histone H1.1 variant (JEDRUSIK and SCHULZE 2001) are required for germline gene silencing. Loss-of-function *mes* mutations or H1.1 RNAi leads to de-silencing of transgenes in the germline. Perhaps long term, continual passage through each germline leads to the progressive removal and/or replacement of histone variants, resulting in an activated or de-activated state of the transgene array. The extent to which histone variant replacement occurs in *C. elegans* spermatogenesis is not known, but it is a common phenomenon that has been found in many organisms (HENNIG 2003).

While imprinting serves a developmentally important role for certain groups of organisms such as mammals, plants, and certain insects, there are other organisms that

have clearly demonstrated the ability to imprint parental DNA, yet for which imprinting has not been found to be developmentally essential. Viable animals having both copies of one or more chromosomes from only one parent have been generated in Drosophila (FUYAMA 1984; KOMMA and ENDOW 1995; MULLER 1958), zebra fish (CORLEY-SMITH *et al*. 1996; STREISINGER *et al*. 1981), and *C. elegans* (HAACK and HODGKIN 1991). Parthenogenesis (i.e ATCHLEY 1977) and androgenesis (i.e. MCKONE and HALPERN 2003) can occur naturally in certain animal and plant species. Clearly, the ability to imprint DNA does not necessitate its use in development. Since the conception of the parental conflict theory (MOORE and HAIG 1991) to explain the evolutionary significance of genomic imprinting in mammals, numerous other hypotheses have emerged that attempt to explain the origin and evolution of imprinting in a broader scope to include non-mammalian systems and in systems that can imprint DNA but where imprinting apparently is not essential (DE LA CASA-ESPERON and SAPIENZA 2003; HAIG and TRIVERS 1995; HURST 1997; LLOYD 2000; MCGOWAN and MARTIN 1997; WALTER and PAULSEN 2003a).

We do not know the extent to which parent-of-origin affects the expression of endogenous *C. elegans* genes. Certainly a strong argument can be made that such effects are either subtle or rare. Three decades of genetic experiments with semi-dominant genetic markers in *C. elegans* have failed to yield any known examples where the parent-of-origin for a particular locus affects its expression. Although many of these genetic studies of semi-dominant loci were not directed toward finding such effects, they would certainly have been detected if such an effect were universal. We do note, however, that certain semi-dominant mutations in *C. elegans* show wide expressivity

among individuals of the same mutant phenotype. It is conceivable that if endogenous genes were subject to subtle imprinting effects, it would manifest itself as differences in expressivity that can be attributed to parent-of-origin of the mutant allele.

A second approach to looking for imprinting in *C. elegans* has been to generate diploid animals in which both copies of a given chromosome are derived from the same parent (HAACK and HODGKIN 1991), either from the male parent or hermaphrodite parent. Such animals have been produced for the X chromosome and each of the five autosomes. In each case the resulting animals can be viable and fertile. Although these experiments rule out an essential role for imprinting of any single locus or chromosome in *C. elegans*, the experiments do not rule out subtle effects on phenotype or (even more significantly) quantitative effects on viability.

A third approach to detecting parent-of-origin effects on native genes in *C. elegans* has been taken by the lab of Bill Kelly (BEAN *et al.* 2004). They have examined the overall modification state of the X chromosome during gametogenesis and just after fertilization. Strikingly, the spermatogenesis-derived X chromosome (in both males and hermaphrodites) shows a strong heterochromatin-like imprint (i.e. lack of dimethylated H3-Lys4 and di-acetylated H3-Lys9/Lys14). Although the differential modification of the X chromosome in response to genetic history (as observed by Bean *et al.*) has some similarity with the imprinting-modulated gene expression that we observe, we note several differences: First, we observe at most a modest effect of parental sex and pairing state relative to a much more substantial difference reflecting gamete-of-origin (sperm versus oocyte). This is distinct from the X chromosome pairing dependence observed by Bean and colleagues. Second, the transgene imprinting that we observe

appears distinct in terms of timing. In particular, the X chromosome imprint disappears from the embryo by the 20-cell stage, and hence might not result in any transcriptional difference between the paternally-derived and maternally-derived chromosomes. In our case, the parent-of-origin effects that we observe are active much later in the life of the animal (*unc-54* is not expressed until just prior to the final division of the myogenic precursors at a stage with several hundred cells) and thus represents a transcriptional difference between maternally versus paternally-derived chromosomes. The timing could indicate a fundamentally different mechanism (i.e. a dependence on independent chromatin modifications), or a common mechanism that derives from some relatively stable chromatin modification that is distinct from those tested with the antibody probes that have been used by Bean and colleagues to date.

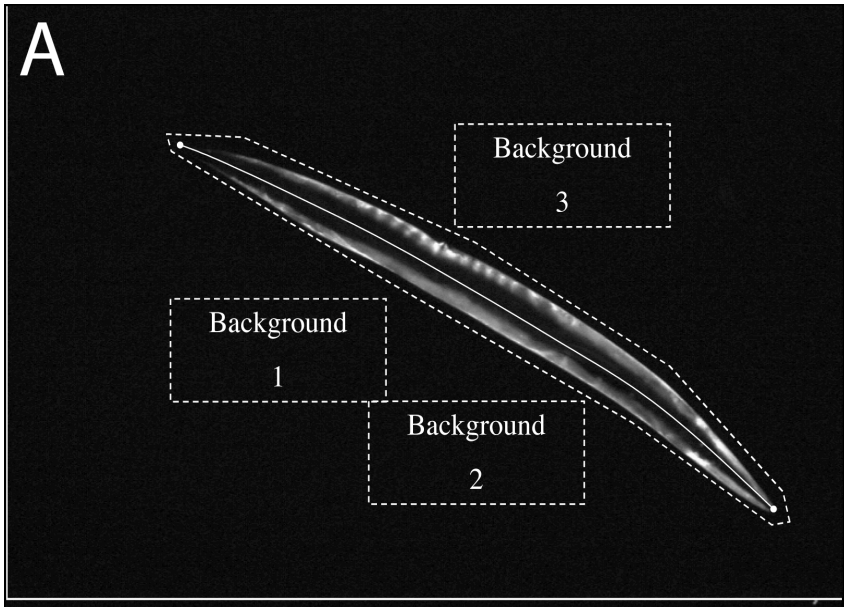
In conclusion, we have developed a quantitative assay to measure average GFP expression of *C. elegans* populations. We have found that among genotypically identical animals, those that receive an *unc-54::gfp* transgene from sperm show, on average, a reproducible greater expression of the transgene compared to animals that receive the same transgene from oocyte. Moreover, when the transgene was kept in the same gamete lineage for multiple, consecutive generations, the transgene acquired a more stable imprint that required multi-generation passage through the opposite germline to completely reset. Since the parent-of-origin effect was observed for multiple *unc-54::gfp* transgenes and for a non-*unc-54* GFP transgene, the ability of *C. elegans* to imprint DNA may not be limited to any unique transgene or integration site. Furthermore, the fact that hermaphrodite sperm and male sperm express an imprinted transgene equivalently

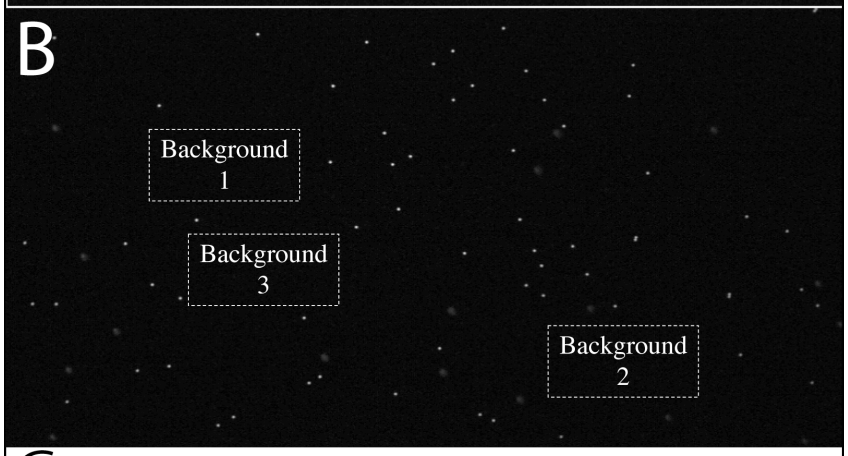
suggests that the source of the imprint may be the gamete type and not the physiological sex of the animal.

FIGURES

Figure 3.1. GFP quantitation assay using NIH Image. (A) Images of experimental, PD4251, and N2 animals to be quantified were captured into NIH Image v1.6.3 followed by fluorescence measurements as described in part C. (B) Fluorescent beads used as standard. Average fluorescence of three background areas was determined for a given fluorescent bead image. This background fluorescence was subtracted from the gross fluorescence signal of each of 10-30 beads (sampled randomly), followed by determination of the average fluorescence of the sampled beads. (C) Method of determining net, normalized signal for PD4251 and experimental animals. For each sample population whose fluorescence measurements was to be determined, we captured four sets of images: one set of the experimental animals, one image of a fluorescent bead standard, one set of PD4251 L4 animals, and one set of 5-10 N2 animals (autofluorescence control). For each N2 animal, net signal was determined by subtracting the average of three background values, multiplying by the area of the animal, and then dividing by the length of the animal. An average N2 value was then determined for each experiment. (We always used N2 animals of the same developmental stage as the experimental animal population. For example, for an experimental population consisting of L4 animals, we determined an N2 average using N2 animals at the L4 stage). For each experimental animal, the net signal ($x_{e,i}$) was determined by subtracting the average of three background values plus the average N2 fluorescence, multiplying by the area ($A_{e,i}$) of the animal, and dividing by the length of the animal ($l_{e,i}$). The net, normalized signal of each animal ($x_{e,i,(normalized)}$) is the ratio between $x_{e,i}$ and average of fluorescent beads. The same procedure was applied to determine fluorescence intensities of PD4251 animals.

Figure 3.1





C

average N2 autofluorescence

$$\bar{x}_{\mathrm{N2}} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\left(I_{\mathrm{N2},i} - B_{\mathrm{N2}}\right)\cdot A_{\mathrm{N2},i}}{l_{\mathrm{N2},i}}\right]$$

average of beads

$$\bar{x}_{\mathrm{bead}} = \frac{1}{n}\sum_{i=1}^{n}\left(I_{\mathrm{bead},i} - B_{\mathrm{bead}}\right)$$

n = number of animals
l = length
A = area
I = gross fluorescence
B = avg. background fluorescence

net signal (non-normalized)

$$x_{\mathrm{e},i} = \frac{\left[I_{\mathrm{e},i} - \left(\bar{x}_{\mathrm{N2}} + B_{\mathrm{e}}\right)\right]\cdot A_{\mathrm{e},i}}{l_{\mathrm{e},i}}$$

normalized net signal

$$x_{\mathrm{e},i(\mathrm{normalized})} = \frac{x_{\mathrm{e},i}}{\bar{x}_{\mathrm{bead}}}$$

Figure 3.2. GFP fluorescence of PD3815 and three of its integrated derivatives. PD3872 is PD3852 out-crossed once. All pictures were taken at the same sitting and under identical settings. Animals were alive at observation.

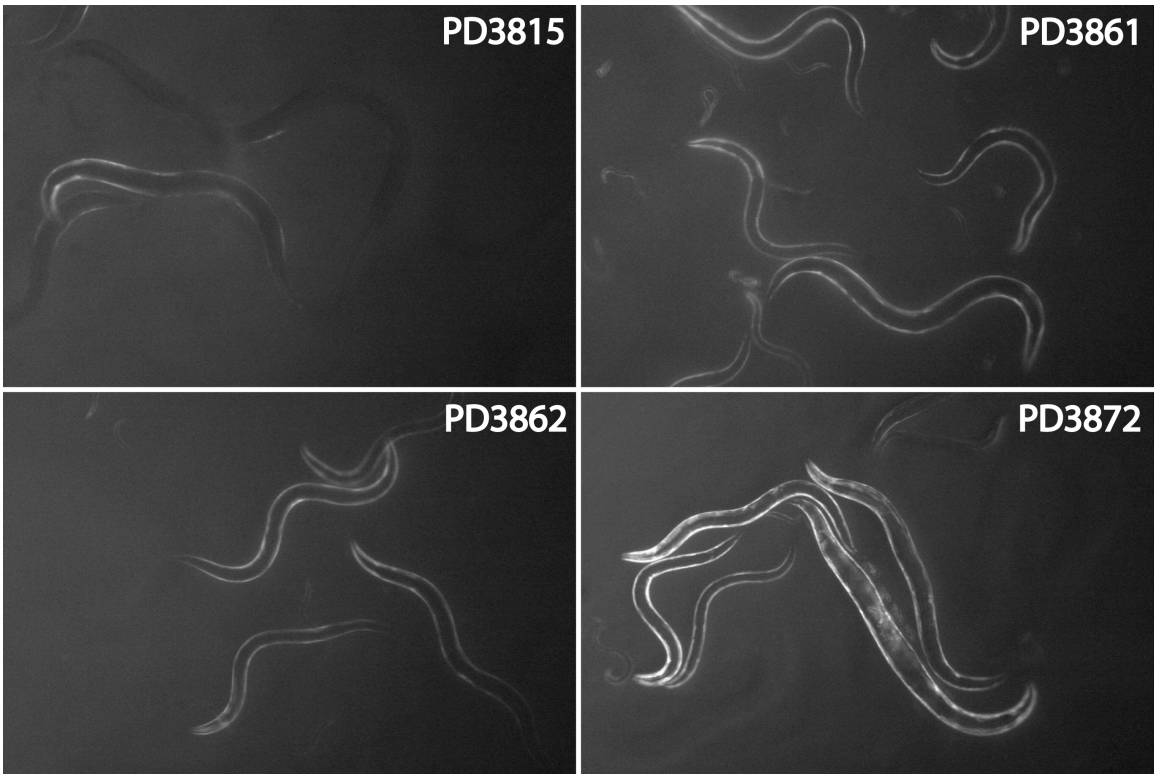


Figure 3.3. Analysis of transgene array copy-number in *ccEx3815* and its various integrated derivatives. (A) Genomic double-digests (*Mfe I* + *PspOM I*) of *ccEx3815*, *ccIn3852*, *ccIn3861*, and *ccIn3862* in the various lines that carry them. Insertion *ccIs9385* (FIRE *et al*. 1998a) carries an integrated *unc-54*::*gfp* tandem array. It was established using *unc-54::gfp*-containing plasmids similar to those used to establish *ccEx3815*: pPD96.02(*unc-54*::*gfp-lacZ*, nuclear localized) and pPD105.19 (*unc-54*::*gfp*, mitochondrial localized). However, the transformation marker for *ccIs9385* is pRF4 (*rol-6*[*su1006*]); whereas it is pC1[*pha-1*(+)] for *ccEx3815*. The arrowhead points to the *pha-1*(+) fragment from pC1. *edIs6*(PD2169) is a non-related *unc-119:gfp* translation fusion. (B) Southern blot analysis of array copy-number in *ccEx3815* and its various derivatives. A 222bp fragment from the *unc-54* promoter was used to probe the gel in part A. The arrowhead points to the N2 and PD2169 *unc-54* bands (each containing two copies of *unc-54*). PD3924 is a marked derivative of PD3862; while PD3872 is PD3852 outcrossed once. Note that *ccEx3815* and all its integrated derivatives appear to have the identical structure; whereas *ccIs9385* is very different structurally.

Figure 3.3

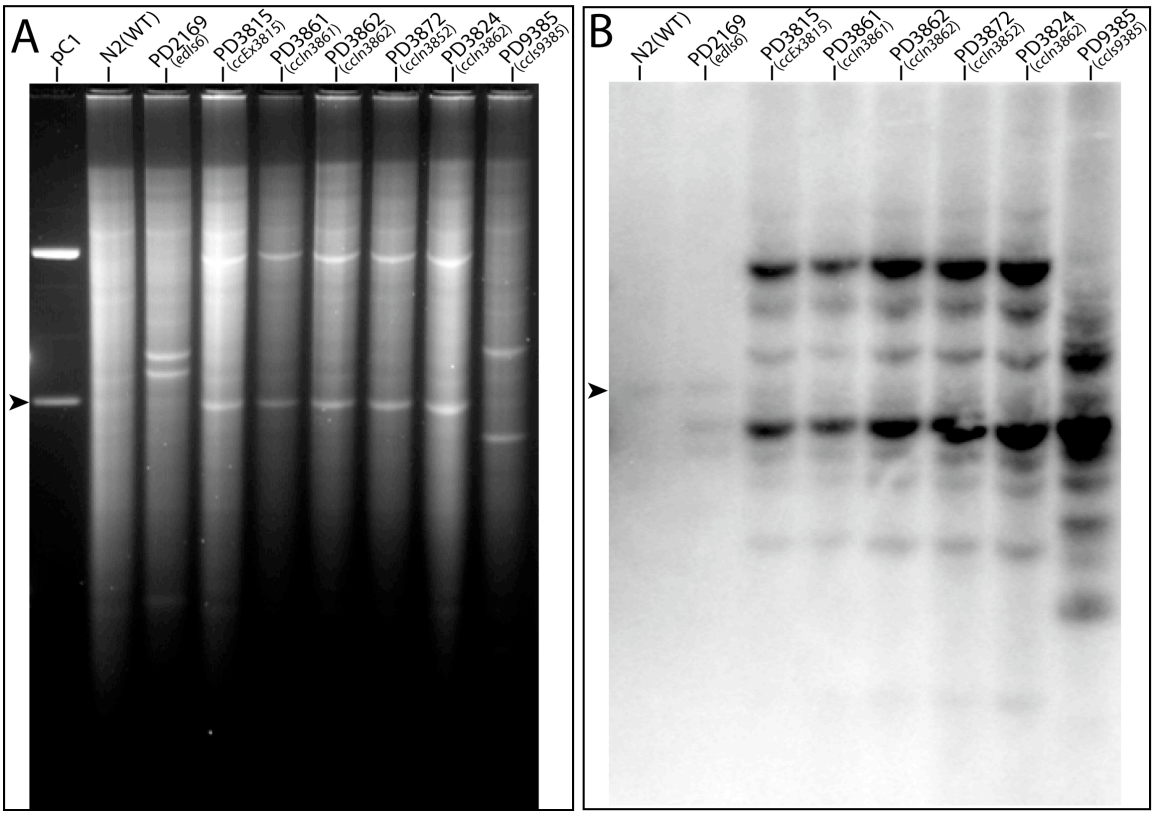
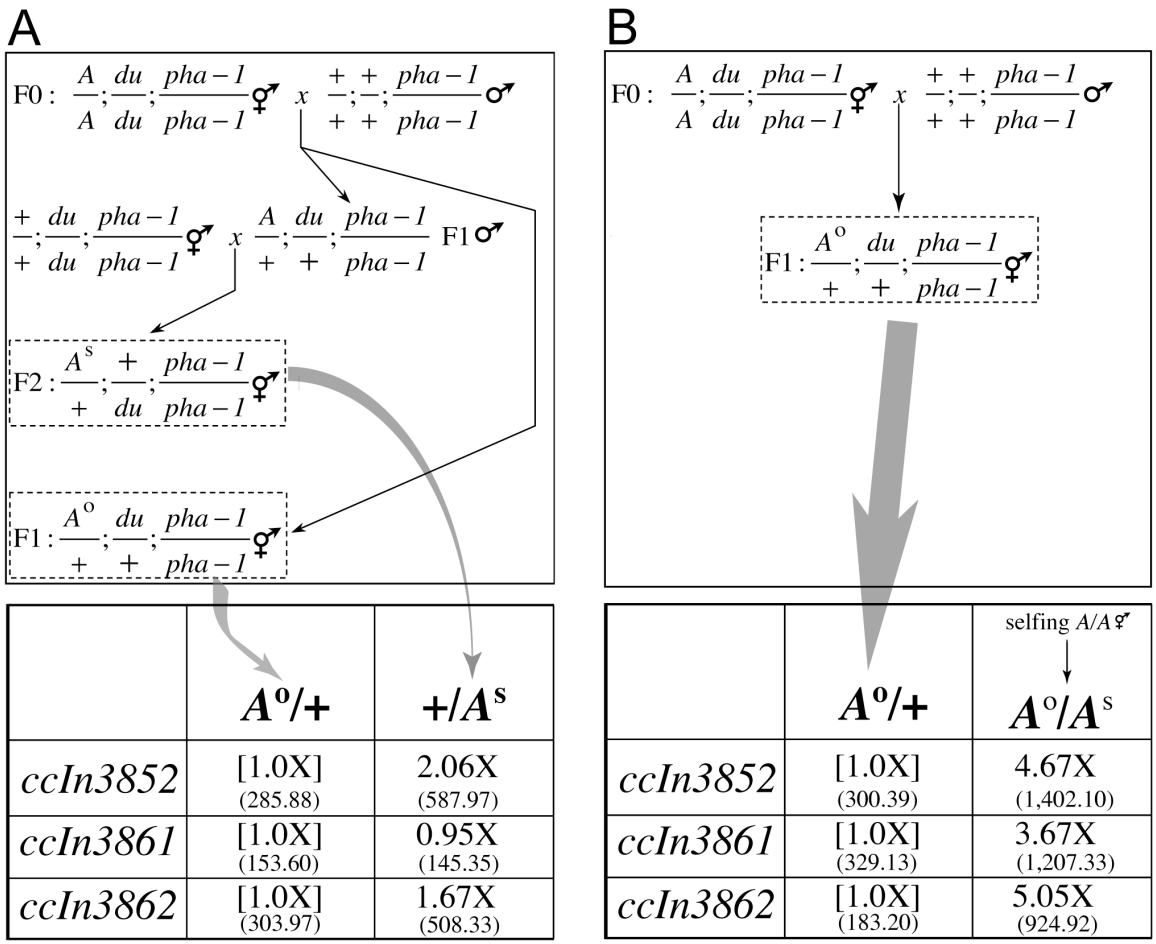

A
pC1
N2(WT)
PD2169 (edIs6)
PD3815 (ccEx3815)
PD3861 (ccIn3861)
PD3862 (ccIn3862)
PD3872 (ccIn3852)
PD3824 (ccIn3862)
PD9385 (ccIs9385)
B
N2(WT)
PD2169 (edIs6)
PD3815 (ccEx3815)
PD3861 (ccIn3861)
PD3862 (ccIn3862)
PD3872 (ccIn3852)
PD3824 (ccIn3862)
PD9385 (ccIs9385)

Figure 3.4. Initial observations of parent-of-origin effect and non-linear expression in *ccIn3852*, *ccIn3861*, and *ccIn3862*. (A) Experimental design used to determine the ratio of A^O/+ to A^S/+ expression in the three transgenic lines. Data are shown both as relative values normalized over A^O/+ (top number) and absolute values (bottom number, in parentheses). Unlike subsequent experiments described in this article in which L4 (larval stage) and adult animals were analyzed as separate data sets, this experiment combined both L4 and adult animals into the same data set. The results for *ccIn3861* were pooled from two independent experiments. The letter "*A*" in the crosses represents the integrated transgene array, and superscripted "o" and "s" indicate that the array was derived from oocyte and sperm, respectively. This convention will be used throughout this article. *du (dpy unc)* are two linked, recessive markers (different for each of the three lines). In all experiments described in this article, only non-*dpy* non-*unc* hermaphrodite cross progeny were used for fluorescence quantitation. Cross progeny animals that were either *dpy* and/or *unc* were never used in any analysis. (B) Experimental design comparing relative expression between array hemizygotes and homozygotes. As in part A, both L4 and adult animals were combined into a single data set.

Figure 3.4



	$A^o/+$	$+/A^s$
ccIn3852	[1.0X] (285.88)	2.06X (587.97)
ccIn3861	[1.0X] (153.60)	0.95X (145.35)
ccIn3862	[1.0X] (303.97)	1.67X (508.33)

	$A^o/+$	selfing A/A ♀ → A^o/A^s
ccIn3852	[1.0X] (300.39)	4.67X (1,402.10)
ccIn3861	[1.0X] (329.13)	3.67X (1,207.33)
ccIn3862	[1.0X] (183.20)	5.05X (924.92)

Figure 3.5. A critical test for parent-of-origin effects in *ccIn3862*. (A) Twelve independent experiments were organized into a matrix. The matrix shows parental genotypes, parental gamete contributions, and progeny genotypes. Red, blue, or green indicates that the transgene array was derived from oocyte, sperm, or both, respectively. (B) Crosses from which animals used for GFP quantitation were derived. In many cases, multi-step crosses were required to arrive at the cross shown in each box.
[*u* = *unc-13*(*e1091*), *d* = *dpy-5*(*e61*), *du* = *dpy-5*(*e61*) *unc-13*(*e1091*)]

Figure 3.5

A

Transmission from oocyte: genotype →	Transmission from oocyte: gamete →	Transmission from sperm: +/+	Transmission from sperm: +/A^s	Transmission from sperm: A^s/+	Transmission from sperm: A^s/A^s
		+	+	A^s	A^s
+/+	+	+/+	+/+	+/A^s	+/A^s
+/A^o	+	+/+	+/+	+/A^s	+/A^s
A^o/+	A^o	A^o/+	A^o/+	A^o / A^s	A^o / A^s
A^o/A^o	A^o	A^o/+	A^o/+	A^o / A^s	A^o / A^s

B

Transmission from oocyte: genotype →	Transmission from oocyte: gamete →	Transmission from sperm: +/+	Transmission from sperm: +/A^s	Transmission from sperm: A^s/+	Transmission from sperm: A^s/A^s
		+	+	A^s	A^s
+/+	+	+/+	+/+	$\frac{+}{+} \times \frac{u}{A\ d}$	$\frac{+}{+} \times \frac{A\ u}{A\ d}$
+/A^o	+	+/+	+/+	$\frac{A\ du}{u} \times \frac{u}{A\ d}$	$\frac{A\ du}{u} \times \frac{A\ u}{A\ d}$
A^o/+	A^o	$\frac{A\ u}{du} \times \frac{+}{+}$	$\frac{A\ u}{du} \times \frac{+}{A\ u}$	$\frac{A\ u}{du} \times \frac{u}{A\ d}$	$\frac{A\ u}{du} \times \frac{A\ u}{A\ d}$
A^o/A^o	A^o	$\frac{A\ u}{A\ u} \times \frac{+}{+}$	$\frac{A\ u}{A\ u} \times \frac{+}{A\ u}$	$\frac{A\ u}{A\ u} \times \frac{u}{A\ d}$	$\frac{A\ u}{A\ u} \times \frac{A\ u}{A\ d}$

Figure 3.6. Results of the matrix experiment. (A,B) For each experiment, we separated the adult and L4 data sets in our analyses because the two stages showed a significant difference in the level of GFP expression, and we were concerned that treating both stages as a single data set would confound our analysis. Each number in the matrix is a population average after normalization over the fluorescent bead standard. Color schemes are as indicated for Figure 3.5. (C) Statistical analyses of the matrix data. The t-test column shows *p* values in scientific notation (i.e. 2.65E-07 = 2.65×10^{-7}). We considered $p<0.05$ to be statistically significant. Experiments C4 and D3 shared the same set of bead and PD4251 images. [*n* = sample size, s.d. = standard deviation, s.e.m. = standard error of the mean]

Figure 3.6

A

L4			A	B	C	D
			+/+	+/A^s	A^s/+	A^s/A^s
			+	+	A^s	A^s
1	+/+	+	+/+	+/+	6.42	6.92
2	+/A^o	+	+/+	+/+	6.02	6.52
3	A^o/+	A^o	4.22	2.78	13.47	15.67
4	A^o/A^o	A^o	3.42	3.18	11.90	17.61

B

Adult			A	B	C	D
			+/+	+/A^s	A^s/+	A^s/A^s
			+	+	A^s	A^s
1	+/+	+	+/+	+/+	11.41	17.80
2	+/A^o	+	+/+	+/+	9.83	16.98
3	A^o/+	A^o	8.01	7.62	24.42	33.45
4	A^o/A^o	A^o	7.63	11.08	28.68	35.35

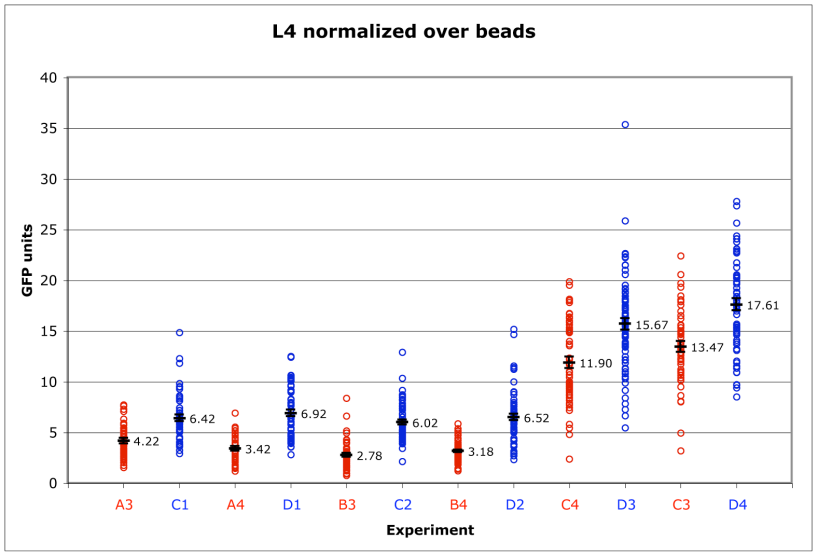
C

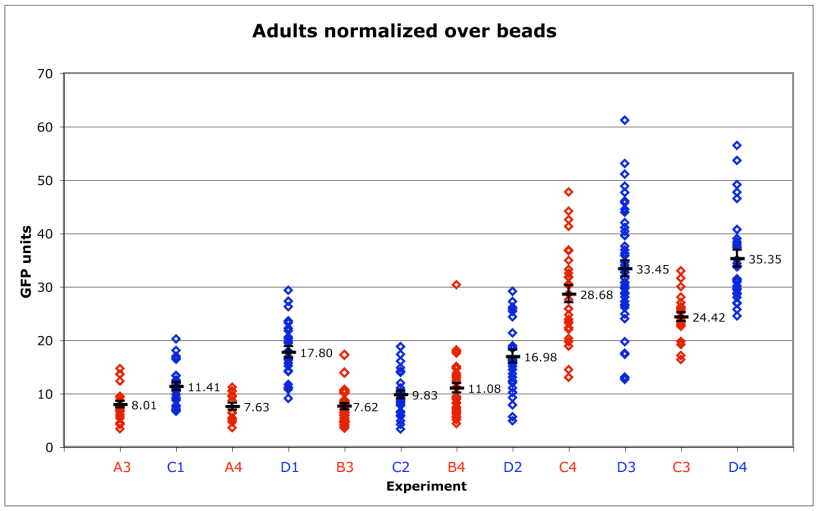
Experimental animals normalized over bead		L4					adults				
	Exp	n	mean	s.d.	s.e.m.	t-test	n	mean	s.d.	s.e.m.	t-test
	A3	49	4.22	1.60	0.23	2.65E-07	25	8.01	3.04	0.61	1.09E-03
	C1	52	6.42	2.34	0.32		25	11.41	3.81	0.76	
	A4	35	3.42	1.39	0.23	1.41E-13	15	7.63	2.50	0.65	9.70E-10
	D1	58	6.92	2.49	0.33		26	17.80	5.51	1.08	
	B3	50	2.78	1.39	0.20	1.81E-17	26	7.62	3.12	0.61	2.20E-02
	C2	58	6.02	1.88	0.25		29	9.83	3.82	0.71	
	B4	63	3.18	1.00	0.13	1.18E-15	38	11.08	5.27	0.86	1.99E-04
	D2	67	6.52	2.60	0.32		31	16.98	6.74	1.21	
	C4	53	11.90	4.06	0.56	5.94E-06	29	28.68	8.67	1.61	3.30E-02
	D3	71	15.67	4.78	0.57		49	33.45	10.42	1.49	
	C3	50	13.47	3.84	0.54	1.24E-06	24	24.42	4.01	0.82	2.10E-07
	D4	62	17.61	4.70	0.60		29	35.35	8.40	1.56	

PD4251(L4) normalized over bead	Exp	L4 n	mean	s.d.	s.e.m.	t-test
	A3	19	3.96	3.58	0.82	1.90E-01
	C1	18	4.18	6.07	1.43	
	A4	15	4.58	7.25	1.87	7.16E-01
	D1	17	4.68	7.86	1.91	
	B3	20	3.81	3.34	0.75	1.68E-01
	C2	22	3.97	4.13	0.88	
	B4	22	3.62	3.94	0.84	5.49E-01
	D2	20	3.55	3.86	0.86	
	C4	18	3.69	4.67	1.10	1.00E+00
	D3	18	3.69	4.67	1.10	
	C3	20	3.67	4.97	1.11	9.93E-02
	D4	20	3.94	5.24	1.17	

Figure 3.7. Scatter plots of data sets from the matrix experiments. Diametrically opposite pairs in the matrix (i.e. A3 versus C1) are grouped next to each other. Numbers indicate the population average for that set. Error bars indicate standard error of the mean. Note each graph has a different vertical scale. Red or blue indicates that the transgene array was oocyte or sperm-derived, respectively.

Figure 3.7


L4 normalized over beads
GFP units
Experiment
A3 C1 A4 D1 B3 C2 B4 D2 C4 D3 C3 D4
4.22 6.42 3.42 6.92 2.78 6.02 3.18 6.52 11.90 15.67 13.47 17.61


Adults normalized over beads
GFP units
Experiment
A3 C1 A4 D1 B3 C2 B4 D2 C4 D3 C3 D4
8.01 11.41 7.63 17.80 7.62 9.83 11.08 16.98 28.68 33.45 24.42 35.35

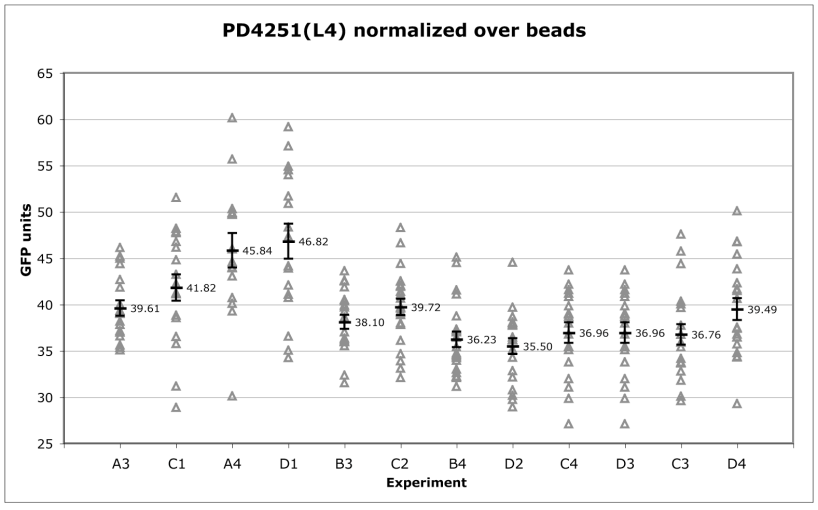
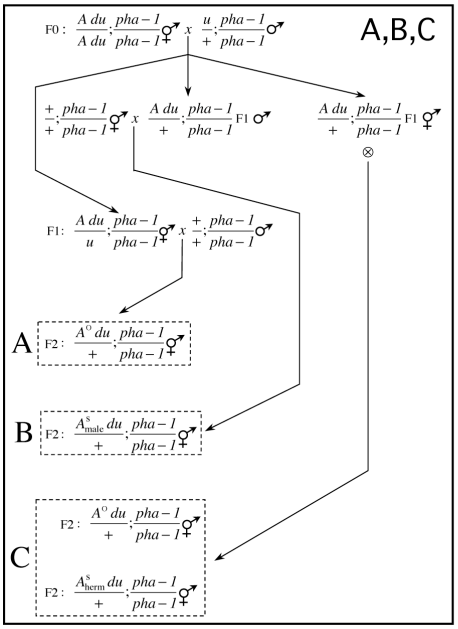

PD4251(L4) normalized over beads
GFP units
Experiment
A3 C1 A4 D1 B3 C2 B4 D2 C4 D3 C3 D4
39.61 41.82 45.84 46.82 38.10 39.72 36.23 35.50 36.96 36.96 36.76 39.49

Figure 3.8. Comparison of *ccIn3862* expression from male sperm versus hermaphrodite sperm. (A-D) Experimental design and the resulting data analysis. Images of animals from all three crosses A, B, and C were captured during the same sitting and therefore shared the same set of fluorescent bead and PD4251 images. Only L4 animals were used in the analysis. (E) Histograms depicting the distribution of *ccIn3862* expression in each data set. The vertical line in each graph indicates the population average for that data set. [*u* = *unc-13*(*e1091*), *d* = *dpy-5*(*e61*), *du* = *dpy-5*(*e61*) *unc-13*(*e1091*)]

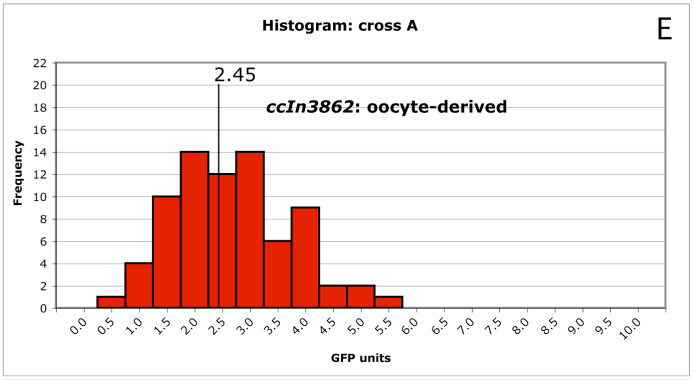
Figure 3.8

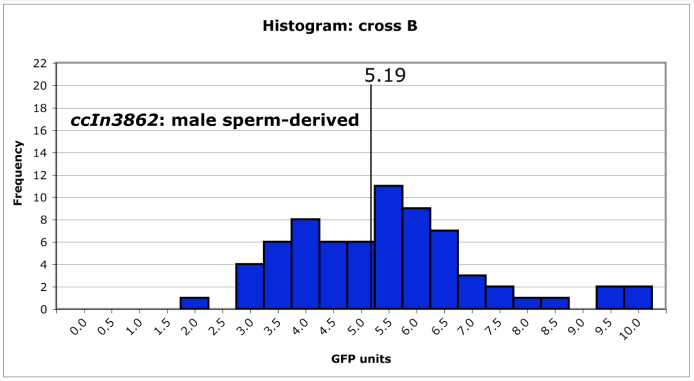


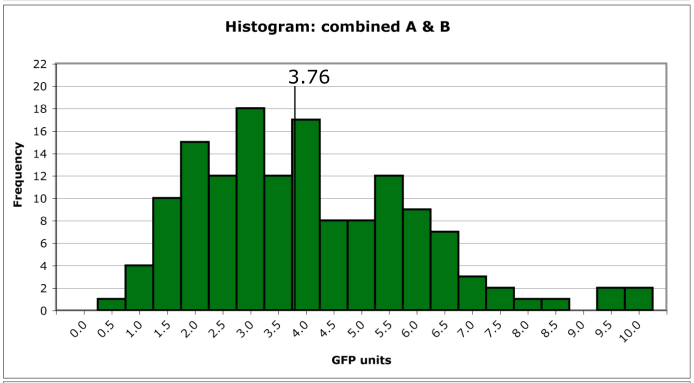
D **Statistics: hermaphrodite vs. male sperm**

cross	n	mean	s.d.	s.e.m.	t-test
A	75	2.45	1.07	0.12	1.29E-20
B	69	5.19	1.70	0.20	
A	75	2.45	1.07	0.12	1.63E-11
C	110	3.90	1.67	0.16	
B	69	5.19	1.70	0.20	1.89E-06
C	110	3.90	1.67	0.16	
AB	144	3.76	1.96	0.16	5.48E-01
C	110	3.90	1.67	0.16	

E







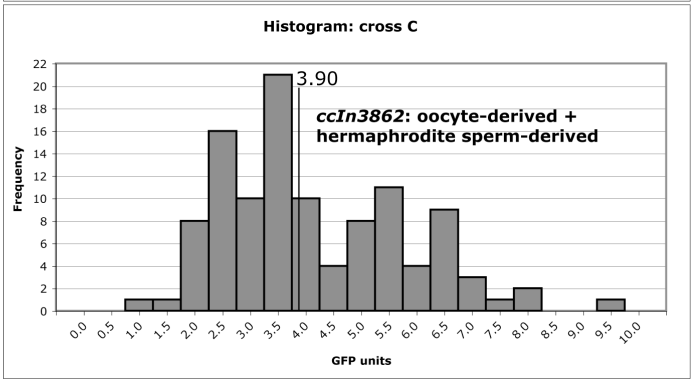


Figure 3.9. A genetic test for pairing effects on *ccIn3862* expression. (A-C) Experimental design to test for pairing effects on *ccIn3862* expression. (D) Analysis for resulting data. Only the L4 data set is shown. The adult data set (data not shown) exhibits a similar trend as the L4 data set. (E) Histograms depicting distribution of *ccIn3862* expression from each data set. The vertical line in each graph indicates the population average for that data set. [*du* = *dpy-17*(*e164*) *unc-69*(*e587*)]

Figure 3.9

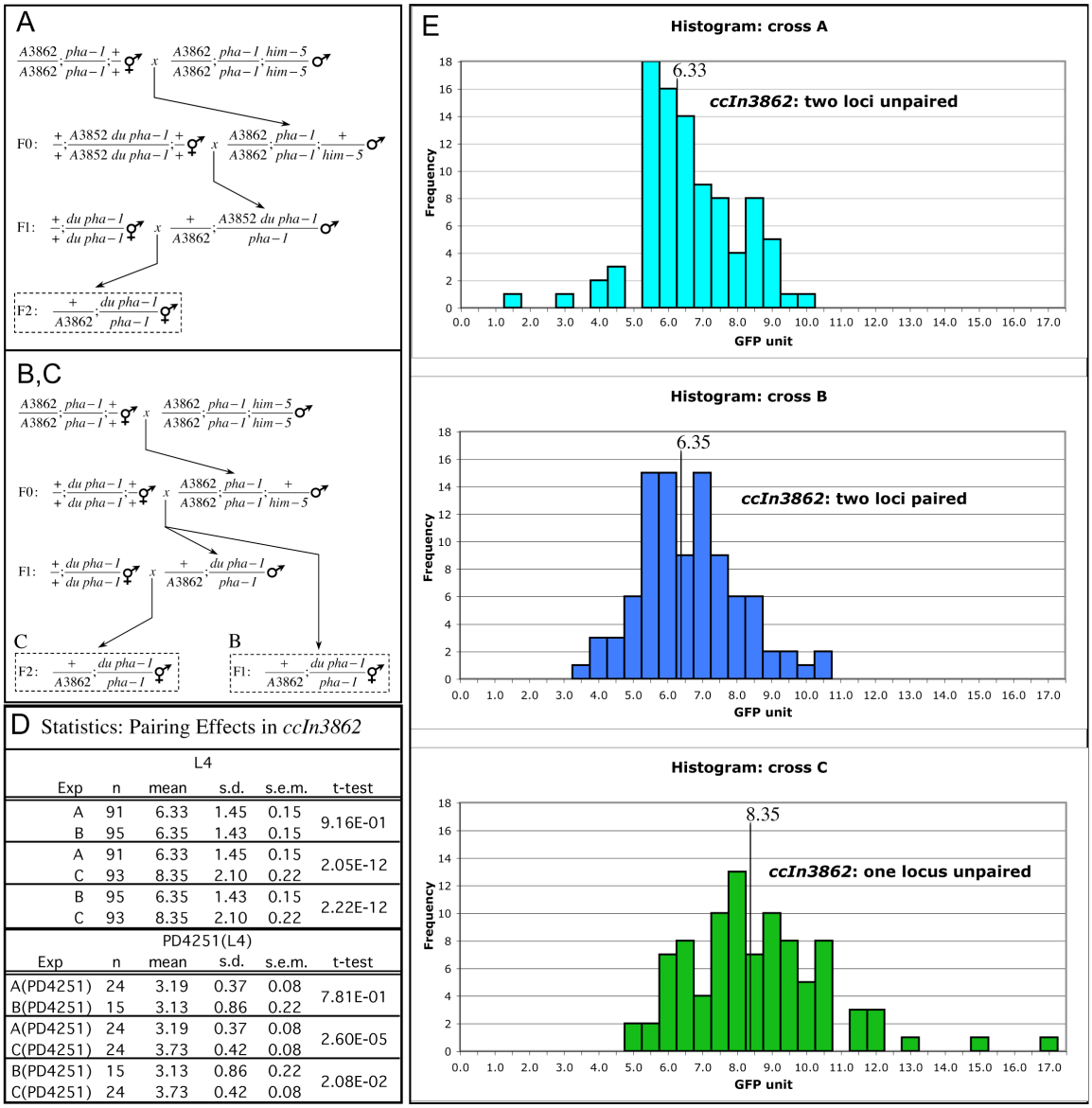

A
F0:
F1:
F2:
B,C
F0:
F1:
C
F2:
B
F1:
D Statistics: Pairing Effects in ccIn3862
L4
Exp n mean s.d. s.e.m. t-test
A 91 6.33 1.45 0.15
B 95 6.35 1.43 0.15 9.16E-01
A 91 6.33 1.45 0.15
C 93 8.35 2.10 0.22 2.05E-12
B 95 6.35 1.43 0.15
C 93 8.35 2.10 0.22 2.22E-12
PD4251(L4)
Exp n mean s.d. s.e.m. t-test
A(PD4251) 24 3.19 0.37 0.08
B(PD4251) 15 3.13 0.86 0.22 7.81E-01
A(PD4251) 24 3.19 0.37 0.08
C(PD4251) 24 3.73 0.42 0.08 2.60E-05
B(PD4251) 15 3.13 0.86 0.22
C(PD4251) 24 3.73 0.42 0.08 2.08E-02
E
Histogram: cross A
6.33
ccIn3862: two loci unpaired
Frequency
GFP unit
Histogram: cross B
6.35
ccIn3862: two loci paired
Frequency
GFP unit
Histogram: cross C
8.35
ccIn3862: one locus unpaired
Frequency
GFP unit

Figure 3.10. Consequences of long term uni-gametic maintenance of *ccIn3862*. (A) Schematic diagram showing long term maintenance of *ccIn3862* through each gamete lineage. The transgene array was transmitted as a hemizygote through the oocyte (red) from F2-F12, switched to sperm from F13-F23 (blue), switched back to oocyte from F24-F33, then back to sperm from F34-F44. From F45-F52, the array was alternately transmitted through each germline for one generation. At F12, a parallel experiment was splintered off from the main experiment in which the transgene array was kept continuously in the oocyte till F34 (solid red arrow). A similar parallel experiment was performed for sperm transmission (solid blue arrow). (B) Actual genetic crosses as depicted by the diagram in (A). Assayed animals are indicated in dashed rectangular boxes. Double parallel bars (||) over dashed arrows indicate generations that are not shown in the diagram. For every mating throughout the experiment, progeny classes were carefully noted to ensure that recombination had not occurred between the transgene array and the two recessive markers. [*Adu* = *ccIn3862 dpy-5*(*e61*) *unc-13*(*e1091*), *du* = *dpy-5*(*e61*) *unc-13*(*e1091*)]. (C,D,E) Results of the germline maintenance experiment displayed as a scatter plot. The main experiment is shown in the middle graph. Red indicates that the transgene array was oocyte-derived; while blue indicates that it was sperm-derived. Double parallel bars (||) between two data groups indicate non-consecutive generations. Left of the long, dashed vertical line, note that the horizontal axis contains skipped generations. Note that the vertical scale is different for each graph. (F) Statistical analysis of the germline maintenance experiment. Color schemes are as indicated in CDE.

Figure 3.10

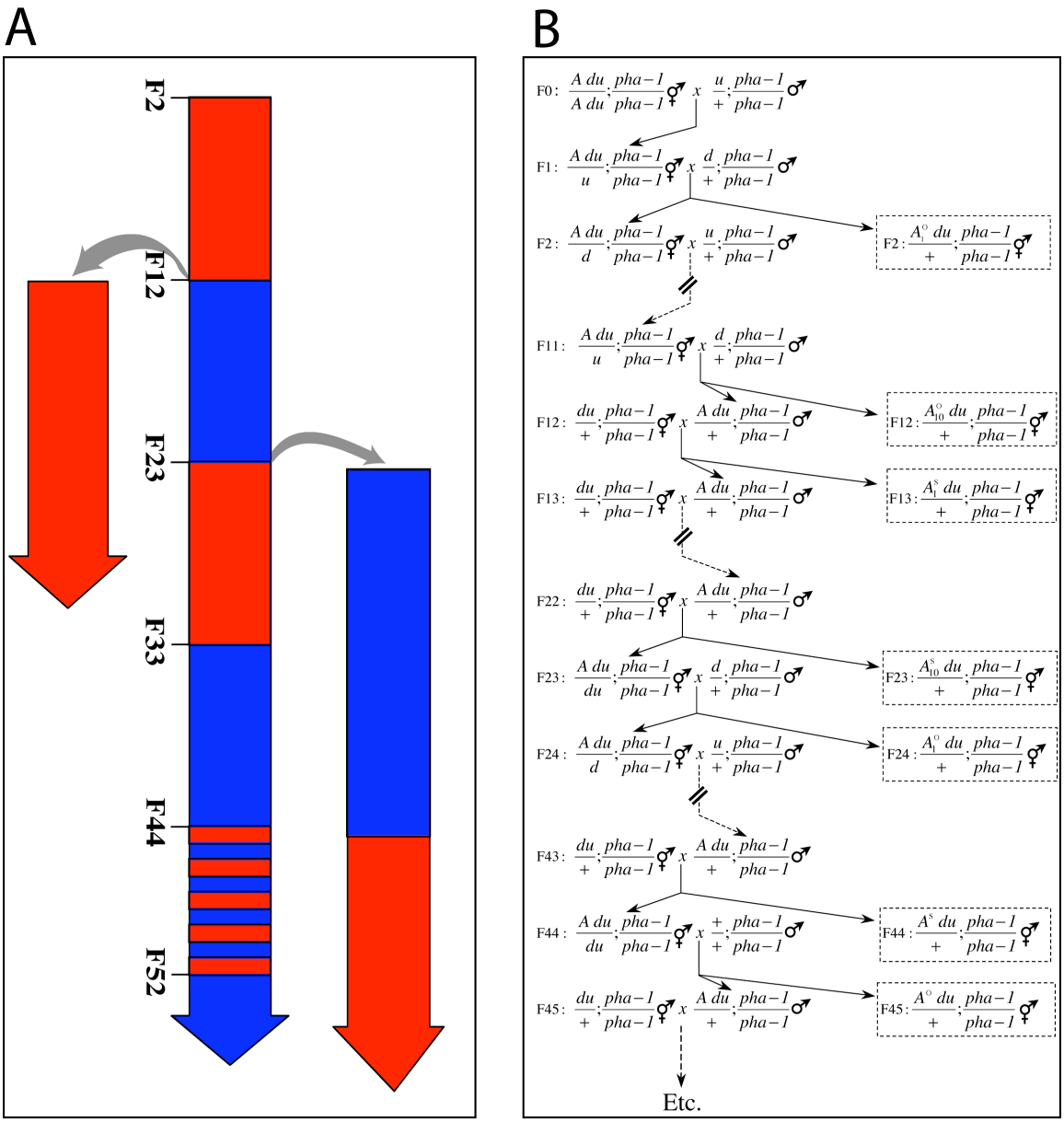

A
B
F2
F12
F23
F33
F44
F52
F0 : A du / A du ; pha−1 / pha−1 ♂ x u / + ; pha−1 / pha−1 ♂
F1 : A du / u ; pha−1 / pha−1 ♀ x d / + ; pha−1 / pha−1 ♂
F2 : A du / d ; pha−1 / pha−1 ♀ x u / + ; pha−1 / pha−1 ♂
F2 : A1° du / + ; pha−1 / pha−1 ♀
F11 : A du / u ; pha−1 / pha−1 ♀ x d / + ; pha−1 / pha−1 ♂
F12 : du / + ; pha−1 / pha−1 ♀ x A du / + ; pha−1 / pha−1 ♂
F12 : A10° du / + ; pha−1 / pha−1 ♀
F13 : du / + ; pha−1 / pha−1 ♀ x A du / + ; pha−1 / pha−1 ♂
F13 : A1s du / + ; pha−1 / pha−1 ♀
F22 : du / + ; pha−1 / pha−1 ♀ x A du / + ; pha−1 / pha−1 ♂
F23 : A du / du ; pha−1 / pha−1 ♀ x d / + ; pha−1 / pha−1 ♂
F23 : A10s du / + ; pha−1 / pha−1 ♀
F24 : A du / d ; pha−1 / pha−1 ♀ x u / + ; pha−1 / pha−1 ♂
F24 : A1° du / + ; pha−1 / pha−1 ♀
F43 : du / + ; pha−1 / pha−1 ♀ x A du / + ; pha−1 / pha−1 ♂
F44 : A du / du ; pha−1 / pha−1 ♀ x + / + ; pha−1 / pha−1 ♂
F44 : As du / + ; pha−1 / pha−1 ♀
F45 : du / + ; pha−1 / pha−1 ♀ x A du / + ; pha−1 / pha−1 ♂
F45 : A° du / + ; pha−1 / pha−1 ♀
Etc.

Figure 3.10 (*continued*)

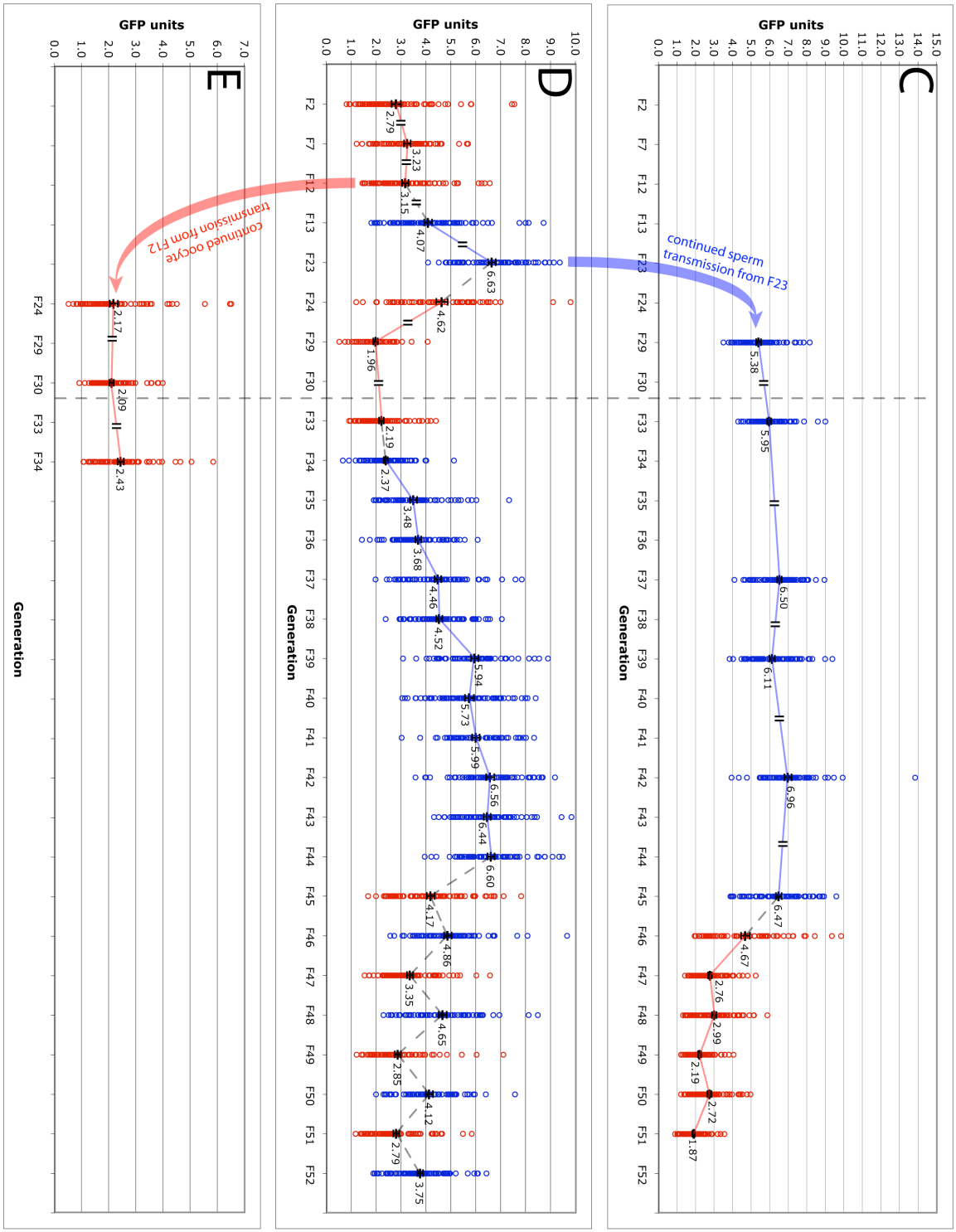


Figure 3.10 (*continued*)

F

Statistics: Figure 9B(top graph)					
Gen.	n	mean	s.d.	s.e.m	t-test
F29	64	5.38	1.04	0.13	9.24E-04
F33	71	5.95	0.90	0.11	7.03E-04
F37	72	6.50	1.01	0.12	4.43E-02
F39	65	6.11	1.21	0.15	5.70E-04
F42	62	6.96	1.49	0.19	4.18E-02
F45	74	6.47	1.31	0.15	4.73E-09
F46	62	4.67	1.87	0.24	1.02E-10
F47	68	2.76	0.82	0.10	1.29E-01
F48	74	2.99	1.00	0.12	1.10E-07
F49	65	2.19	0.65	0.08	8.76E-05
F50	73	2.72	0.89	0.10	6.47E-11
F51	82	1.87	0.52	0.06	

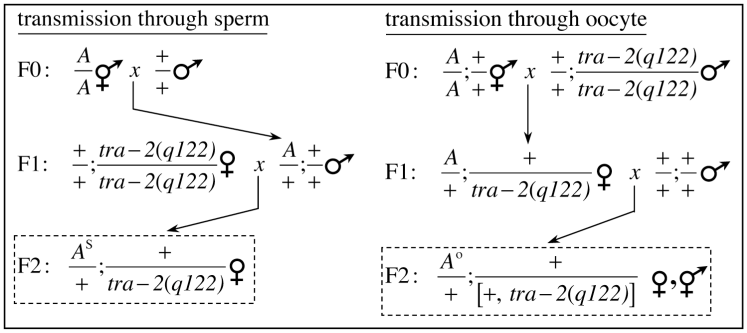
Statistics: Figure 9C(middle graph)					
Gen.	n	mean	s.d.	s.e.m.	t-test
F2	84	2.79	1.68	0.18	5.00E-02
F7	61	3.23	1.02	0.13	6.62E-01
F12	81	3.15	1.18	0.13	1.01E-05
F13	90	4.07	1.45	0.15	1.94E-23
F23	75	6.63	1.34	0.15	2.75E-11
F24	66	4.62	1.82	0.22	4.42E-18
F29	58	1.96	0.68	0.09	6.92E-02
F33	74	2.19	0.80	0.09	1.58E-01
F34	77	2.37	0.74	0.08	6.29E-10
F35	67	3.48	1.15	0.14	2.74E-01
F36	71	3.68	0.98	0.12	1.88E-05
F37	83	4.46	1.21	0.13	7.49E-01
F38	72	4.52	1.01	0.12	5.02E-11
F39	61	5.94	1.22	0.16	3.75E-01
F40	64	5.73	1.46	0.18	2.70E-01
F41	56	5.99	1.13	0.15	1.07E-02
F42	60	6.56	1.27	0.16	5.76E-01
F43	66	6.44	1.12	0.14	4.37E-01
F44	70	6.60	1.15	0.14	1.65E-22
F45	75	4.17	1.35	0.16	1.70E-03
F46	69	4.86	1.25	0.15	1.55E-12
F47	71	3.35	1.04	0.12	2.10E-09
F48	66	4.65	1.30	0.16	3.27E-15
F49	76	2.85	1.04	0.12	3.75E-10
F50	61	4.12	1.11	0.14	1.13E-10
F51	67	2.79	1.00	0.12	3.67E-08
F52	85	3.75	1.01	0.11	

Statistics: Figure 9D(bottom graph)					
Gen.	n	mean	s.d.	s.e.m.	t-test
F24	70	2.17	1.25	0.15	6.33E-01
F30	68	2.09	0.70	0.08	1.71E-02
F34	68	2.43	0.93	0.11	

Figure 3.11. Tests for parent-of-origin effects in non-*unc-54* reporter transgenes. (A) Experimental design. Because these lines were not in the *pha-1*(*e2123ts*) background (and hence the transgenes were not *pha-1(e2123ts)*-selected), we used the strategy shown here. *tra-2*(*q122*) is a dominant mutation that renders hermaphrodites incapable of sperm production, essentially converting *tra-2*(*q122*) hermaphrodites into obligate females (SCHEDL and KIMBLE 1988). *tra-2*(*q122*) males are unaffected. For each transgene tested, both sperm-transmitted and oocyte-transmitted crosses were done synchronously and images of F2 animals resulting from the crosses were captured in the same sitting. This allowed us to directly compare average population fluorescent values from the two crosses without the use of a fluorescent bead standard, by directly normalizing values for A^S/+ relative to A^O/+. (B) Results. Each ratio in the right column is an independent experiment [i.e. two independent experiments for *lin-11*::*gfp*(*nIs106*)]. Asterisks indicate that the experiment was done at 20°C, rather than at 23°C. (C) Statistical analysis for *unc-119*:*gfp*(*edIs6*). Images from all three experiments for *unc-119:gfp*(*edIs6*) were captured using the same camera, but a different microscope was used for Experiment 3.

Figure 3.11

A



B

transgene	A^o/+ : A^s/+
lin-11::*gfp* (*nIs106*)	[1.0X] :1.2X* [1.0X] :1.0X
myo-3::*gfp* (*cc4251*)	[1.0X] :1.1X*
ceh-23::*gfp* (*lqIs27*)	[1.0X] :0.83X* [1.0X] :1.0X* [1.0X] :0.67X (adult)
unc-119:*gfp* (*edIs6*) [translational fusion]	See statistics table for *unc-119* (*edIs6*)
sur-5::*gfp*(*ccEx8438*) [extra-chromosomal]	[1.0X] :1.2X*

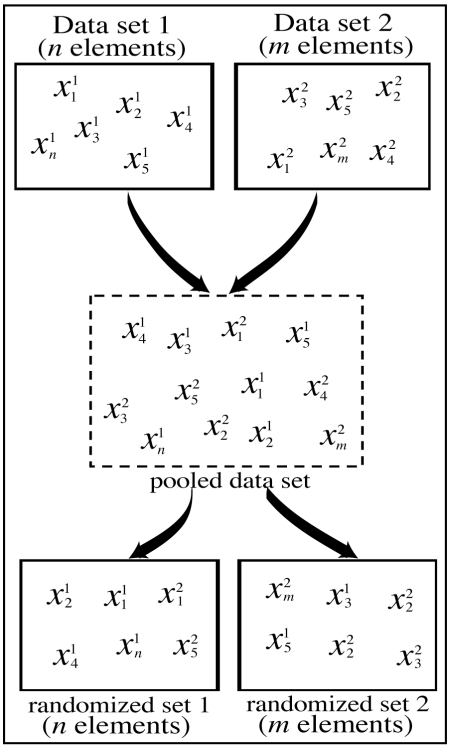
C

statistics: *unc-119*:*gfp*(*edIs6*)

		L4					adult				
Exp.	Genotype	n	mean	s.d.	s.e.m.	t-test	n	mean	s.d.	s.e.m.	t-test
1	A^o/+	16	58.49	12.81	3.20	7.43E-06	18	99.33	27.16	6.40	1.46E-09
	A^s/+	13	92.29	17.34	4.81		19	174.17	28.91	6.63	
2	A^o/+	10	93.26	11.59	3.67	2.20E-03	32	146.00	33.17	5.86	1.78E-03
	A^s/+	9	124.33	21.40	7.13		30	182.37	51.04	9.32	
3	A^o/+	42	215.82	73.78	11.38	4.70E-05	18	285.34	96.92	22.85	8.98E-05
	A^s/+	46	299.63	107.13	15.79		19	586.63	259.23	59.47	

Figure 3.12. *In silico* validation of experimental data. (A) The methodology behind the statistical analysis. (B) Results of the computer simulation. The right four columns show the means of the randomized data sets and the frequency each randomized mean is equal to, greater than, or less than the experimental mean. *Frequencies in the bottom row (AB versus C) do not add up to exactly 10,000 due to rounding errors in computer floating point arithmetic.

A



B

	Experiment	exp. mean	randomized mean	Number of times randomized mean is... = exp. mean	> exp. mean	< exp. mean
matrix	A3(L4)	4.22	5.349	0	10,000	0
	C1(L4)	6.42	5.355	0	0	10,000
	A4(L4)	3.42	5.601	0	10,000	0
	D1(L4)	6.92	5.602	0	0	10,000
	B3(L4)	2.78	4.525	0	10,000	0
	C2(L4)	6.02	4.513	0	0	10,000
	B4(L4)	3.18	4.902	0	10,000	0
	D2(L4)	6.52	4.896	0	0	10,000
	C4(L4)	11.9	14.070	0	9,807	193
	D3(L4)	15.67	14.057	0	193	9,807
	A3(ad)	8.01	9.706	0	9,994	6
	C1(ad)	11.41	9.709	0	6	9,994
	A4(ad)	7.63	14.099	0	10,000	0
	D1(ad)	17.80	14.062	0	0	10,000
	B3(ad)	7.62	8.778	0	9,872	128
	C2(ad)	9.83	8.790	0	128	9,872
	B4(ad)	11.08	13.729	0	10,000	0
	D2(ad)	16.98	13.734	0	0	10,000
	C4(ad)	28.68	31.668	0	9,807	193
	D3(ad)	33.45	31.684	0	193	9,807
sperm equiv.	A	2.45	3.761	0	10,000	0
	B	5.19	3.763	0	0	10,000
	A	2.45	3.313	0	10,000	0
	C	3.90	3.313	0	0	10,000
	B	5.19	4.394	0	0	10,000
	C	3.90	4.397	0	10,000	0
	* AB	3.76	3.821	3	7,246	2,753
	* C	3.90	3.823	3	2,755	7,245

ACKNOWLEDGMENTS

We thank the following people for helpful discussions and/or reagents: A. Villeneuve, V. Corces, M. van Doren, W. Lui, S. Johnson, J. Pak, D. Blanchard, R. Yokoo, S. Kostas, J. Yanowitz, J. Fleenor, E. Whitelaw, B. Kelly, M. Maduro, and S. Cameron. We acknowledge financial support from the National Institutes of Health grants GM37706 and T32GM07231 (K.S.) and institutional support from Stanford University, Carnegie Institution of Washington, and Johns Hopkins University.

REFERENCES

AINSCOUGH, J. F., T. KOIDE, M. TADA, S. BARTON and M. A. SURANI, 1997 Imprinting of *Igf2* and *H19* from a 130 kb YAC transgene. Development **124:** 3621-3632.

ALLEMAN, M., and J. DOCTOR, 2000 Genomic imprinting in plants: observations and evolutionary implications. Plant Mol Biol **43:** 147-161.

ARAMAYO, R., and R. L. METZENBERG, 1996 Meiotic transvection in fungi. Cell **86:** 103-113.

ASSAAD, F. F., K. L. TUCKER and E. R. SIGNER, 1993 Epigenetic repeat-induced gene silencing (RIGS) in Arabidopsis. Plant Mol Biol **22:** 1067-1085.

ATCHLEY, W. R., 1977 Evolutionary consequences of parthogenesis: evidence from the *Warramaba virgo* complex. Proc Natl Acad Sci U S A **74:** 1130-1134.

BEAN, C. J., C. E. SCHANER and W. G. KELLY, 2004 Meiotic pairing and imprinted X chromatin assembly in *Caenorhabditis elegans*. Nat Genet **36:** 100-105.

BESTOR, T. H., 2000 The DNA methyltransferases of mammals. Hum Mol Genet **9:** 2395-2402.

BONGIORNI, S., and G. PRANTERA, 2003 Imprinted facultative heterochromatization in mealybugs. Genetica **117:** 271-279.

BRENNER, S., 1974 The genetics of *Caenorhabditis elegans*. Genetics **77:** 71-94.

BROWN, S. W., 1959 Lecanoid chromosome behavior in three more families of the Coccoidea (Homoptera). Chromosoma **10:** 278-300.

BROWN, S. W., 1961 Fracture and fusion of coccid chromosomes. Nature **191:** 1419-1420.

BROWN, S. W., and W. A. NELSEN-REES, 1961 Radiation analysis of the lecanoid genetic system. Genetics **42:** 510-523.

CAMBARERI, E. B., B. C. JENSEN, E. SCHABTACH and E. U. SELKER, 1989 Repeat-induced G-C to A-T mutations in Neurospora. Science **244:** 1571-1575.

CATTANACH, B. M., and C. V. BEECHEY, 1990 Autosomal and X-chromosome imprinting. Dev Suppl**:** 63-72.

CHAILLET, J. R., T. F. VOGT, D. R. BEIER and P. LEDER, 1991 Parental-specific methylation of an imprinted transgene is established during gametogenesis and progressively changes during embryogenesis. Cell **66:** 77-83.

COHEN, J., 1962 Position-effect variegation at several closely linked loci in *Drosophila melanogaster*. Gerontol Clin (Basel) **47:** 647-659.

CORLEY-SMITH, G. E., C. J. LIM and B. P. BRANDHORST, 1996 Production of androgenetic zebrafish (*Danio rerio*). Genetics **142:** 1265-1276.

CROUSE, H. V., 1960 The controlling element in sex chromosome behavior in Sciara. Genetics **45:** 1429-1443.

DAY, C. D., E. LEE, J. KOBAYASHI, L. D. HOLAPPA, H. ALBERT *et al.*, 2000 Transgene integration into the same chromosome location can produce alleles that express at a predictable level, or alleles that are differentially silenced. Genes Dev **14:** 2869-2880.

DE LA CASA-ESPERON, E., and C. SAPIENZA, 2003 Natural selection and the evolution of genome imprinting. Annu Rev Genet **37:** 349-370.

DELAVAL, K., and R. FEIL, 2004 Epigenetic regulation of mammalian genomic imprinting. Curr Opin Genet Dev **14:** 188-195.

DORER, D. R., and S. HENIKOFF, 1997 Transgene repeat arrays interact with distant heterochromatin and cause silencing in cis and trans. Genetics **147:** 1181-1190.

EPSTEIN, H. F., R. H. WATERSTON and S. BRENNER, 1974 A mutant affecting the heavy chain of myosin in *Caenorhabditis elegans*. J Mol Biol **90:** 291-300.

FEDORIW, A. M., P. STEIN, P. SVOBODA, R. M. SCHULTZ and M. S. BARTOLOMEI, 2004 Transgenic RNAi reveals essential function for CTCF in *H19* gene imprinting. Science **303:** 238-240.

FIRE, A., D. ALBERTSON, S. W. HARRISON and D. G. MOERMAN, 1991 Production of antisense RNA leads to effective and specific inhibition of gene expression in *C. elegans* muscle. Development **113:** 503-514.

FIRE, A., and C. MELLO, 1995 Transformation, pp. 451-482 in *Caenorhabditis elegans: Modern Biological Analysis of an Organism*, edited by H. F. EPSTEIN and D. C. SHAKES. Academic Press, San Diego.

FIRE, A., W. KELLY, M. HSU, S. XU, J. AHNN *et al*., 1998a The Uses of Green Fluorescent Protein in *Caenorhabditis elegans*., pp. 153-168 in *Green Fluorescent Protein: Properties, Applications, and Protocols*., edited by M. C. A. S. KANE. Wiley-Liss, Inc., New York.

FIRE, A., S. XU, M. K. MONTGOMERY, S. A. KOSTAS, S. E. DRIVER *et al*., 1998b Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature **391:** 806-811.

FITZPATRICK, G. V., P. D. SOLOWAY and M. J. HIGGINS, 2002 Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. Nat Genet **32:** 426-431.

FOURNIER, C., Y. GOTO, E. BALLESTAR, K. DELAVAL, A. M. HEVER *et al.*, 2002 Allele-specific histone lysine methylation marks regulatory regions at imprinted mouse genes. Embo J **21:** 6560-6570.

FUYAMA, Y., 1984 Gynogenesis in *Drosophila melanogaster*. Jpn. J. Genetics **59:** 91-96.

GARRICK, D., S. FIERING, D. I. MARTIN and E. WHITELAW, 1998 Repeat-induced gene silencing in mammals. Nat Genet **18:** 56-59.

GODAY, C., and M. R. ESTEBAN, 2001 Chromosome elimination in sciarid flies. Bioessays **23:** 242-250.

GOLIC, K. G., M. M. GOLIC and S. PIMPINELLI, 1998 Imprinted control of gene activity in Drosophila. Curr Biol **8:** 1273-1276.

GRANATO, M., H. SCHNABEL and R. SCHNABEL, 1994 *pha-1*, a selectable marker for gene transfer in *C. elegans*. Nucleic Acids Res **22:** 1762-1763.

GREGORY, R. I., T. E. RANDALL, C. A. JOHNSON, S. KHOSLA, I. HATADA *et al.*, 2001 DNA methylation is linked to deacetylation of histone H3, but not H4, on the imprinted genes *Snrpn* and *U2af1-rs1*. Mol Cell Biol **21:** 5426-5436.

GROSSNIKLAUS, U., J. P. VIELLE-CALZADA, M. A. HOEPPNER and W. B. GAGLIANO, 1998 Maternal control of embryogenesis by *MEDEA*, a polycomb group gene in Arabidopsis. Science **280:** 446-450.

HAACK, H., and J. HODGKIN, 1991 Tests for parental imprinting in the nematode *Caenorhabditis elegans*. Mol Gen Genet **228:** 482-485.

HADCHOUEL, M., H. FARZA, D. SIMON, P. TIOLLAIS and C. POURCEL, 1987 Maternal inhibition of hepatitis B surface antigen gene expression in transgenic mice correlates with de novo methylation. Nature **329:** 454-456.

HAIG, D., and R. TRIVERS, 1995 The evolution of parental imprinting: a review of hypotheses, pp. 17-28 in *Genomic Imprinting: Causes and Consequences*, edited by R. OHLSSON, K. HALL and M. RITZEN. Cambridge Univ. Press, Cambridge.

HENIKOFF, S., and T. D. DREESEN, 1989 Trans-inactivation of the Drosophila *brown* gene: evidence for transcriptional repression and somatic pairing dependence. Proc Natl Acad Sci U S A **86:** 6704-6708.

HENNIG, W., 2003 Chromosomal proteins in the spermatogenesis of Drosophila. Chromosoma **111:** 489-494.

HODGKIN, J., H. R. HORVITZ and S. BRENNER, 1979 Nondisjunction mutants of the nematode *Caenorhabditis elegans*. Genetics **91:** 67-94.

HOWELL, C. Y., T. H. BESTOR, F. DING, K. E. LATHAM, C. MERTINEIT *et al.*, 2001 Genomic imprinting disrupted by a maternal effect mutation in the *Dnmt1* gene. Cell **104:** 829-838.

HSIEH, J., J. LIU, S. A. KOSTAS, C. CHANG, P. W. STERNBERG *et al.*, 1999 The RING finger/B-box factor TAM-1 and a retinoblastoma-like protein LIN-35 modulate context-dependent gene silencing in *Caenorhabditis elegans*. Genes Dev **13:** 2958-2970.

HURST, L. D., 1997 Evolutionary theories of genomic imprinting, pp. 211-237 in *Genomic Imprinting: Frontiers in Molecular Biology*, edited by W. REIK and M. A. SURANI. oxford Univ. Press, Oxford.

JEDRUSIK, M. A., and E. SCHULZE, 2001 A single histone H1 isoform (H1.1) is essential for chromatin silencing and germline development in *Caenorhabditis elegans*. Development **128:** 1069-1080.

JOANIS, V., and V. K. LLOYD, 2002 Genomic imprinting in Drosophila is maintained by the products of *Suppressor of variegation* and *trithorax* group, but not *Polycomb* group, genes. Mol Genet Genomics **268:** 103-112.

KEARNS, M., J. PREIS, M. MCDONALD, C. MORRIS and E. WHITELAW, 2000 Complex patterns of inheritance of an imprinted murine transgene suggest incomplete germline erasure. Nucleic Acids Res **28:** 3301-3309.

KELLY, W. G., and A. FIRE, 1998 Chromatin silencing and the maintenance of a functional germline in *Caenorhabditis elegans*. Development **125:** 2451-2456.

KINOSHITA, T., A. MIURA, Y. CHOI, Y. KINOSHITA, X. CAO *et al.*, 2004 One-way control of *FWA* imprinting in Arabidopsis endosperm by DNA methylation. Science **303:** 521-523.

KOMMA, D. J., and S. A. ENDOW, 1995 Haploidy and androgenesis in Drosophila. Proc Natl Acad Sci U S A **92:** 11884-11888.

L'HERNAULT, S. W., D. C. SHAKES and S. WARD, 1988 Developmental genetics of chromosome I spermatogenesis-defective mutants in the nematode *Caenorhabditis elegans*. Genetics **120:** 435-452.

L'HERNAULT, S. W., 1997 Spermatogenesis, pp. 291-294 in *C. elegans II*, edited by D. L. RIDDLE, T. BLUMENTHAL, B. J. MEYER and J. R. PREIS. Cold Spring Harbor Press, New York.

LAMUNYON, C. W., and S. WARD, 1997 Increased competitiveness of nematode sperm bearing the male X chromosome. Proc Natl Acad Sci U S A **94:** 185-189.

LAMUNYON, C. W., and S. WARD, 1998 Larger sperm outcompete smaller sperm in the nematode *Caenorhabditis elegans*. Proc R Soc Lond B Biol Sci **265:** 1997-2002.

LAU, S., K. JARDINE and M. W. MCBURNEY, 1999 DNA methylation pattern of a tandemly repeated *LacZ* transgene indicates that most copies are silent. Dev Dyn **215:** 126-138.

LEWIS, E. G., 1954 The theory and application of a new method of detecting chromosomal rearrangements in *Drosophila melanogaster*. Am. Nat. **88:** 225-239.

LLOYD, V., 2000 Parental imprinting in Drosophila. Genetica **109:** 35-44.

LLOYD, V. K., D. A. SINCLAIR and T. A. GRIGLIATTI, 1999 Genomic imprinting and position-effect variegation in *Drosophila melanogaster*. Genetics **151:** 1503-1516.

MACKENZIE, J. M., JR., F. SCHACHAT and H. F. EPSTEIN, 1978 Immunocytochemical localization of two myosins within the same muslce cells in *Caenorhabditis elegans*. Cell **15:** 413-419.

MADURO, M., and D. PILGRIM, 1995 Identification and cloning of *unc-119*, a gene expressed in the *Caenorhabditis elegans* nervous system. Genetics **141:** 977-988.

MAGER, J., N. D. MONTGOMERY, F. P. DE VILLENA and T. MAGNUSON, 2003 Genome imprinting regulated by the mouse *Polycomb* group protein Eed. Nat Genet **33:** 502-507.

MAGGERT, K. A., and K. G. GOLIC, 2002 The Y chromosome of *Drosophila melanogaster* exhibits chromosome-wide imprinting. Genetics **162:** 1245-1258.

MANN, J. R., and R. H. LOVELL-BADGE, 1987 The development of XO gynogenetic mouse embryos. Development **99:** 411-416.

MARTIN, C. C., and R. A. MCGOWAN, 1995 Parent-of-origin specific effects on the methylation of a transgene in zebrafish. Devel. Genet. **17:** 233-239.

MATZKE, A. J., F. NEUHUBER, Y. D. PARK, P. F. AMBROS and M. A. MATZKE, 1994 Homology-dependent gene silencing in transgenic plants: epistatic silencing loci contain multiple copies of methylated transgenes. Mol Gen Genet **244:** 219-229.

MCGOWAN, R. A., and C. C. MARTIN, 1997 DNA methylation and genome imprinting in the zebrafish, *Danio rerio*: some evolutionary ramifications. Biochem Cell Biol **75:** 499-506.

MCKONE, M. J., and S. L. HALPERN, 2003 The evolution of androgenesis. Am Nat **161:** 641-656.

MELLO, C. C., J. M. KRAMER, D. STINCHCOMB and V. AMBROS, 1991 Efficient gene transfer in *C.elegans*: extrachromosomal maintenance and integration of transforming sequences. Embo J **10:** 3959-3970.

MOORE, T., and D. HAIG, 1991 Genomic imprinting in mammalian development: a parental tug-of-war. Trends Genet **7:** 45-49.

MULLER, H. J., 1958 An androgenic homozygous male. Dros. Info. Service **61:** 215.

OTTE, A. P., and T. H. KWAKS, 2003 Gene repression by *Polycomb* group protein complexes: a distinct complex for every occasion? Curr Opin Genet Dev **13:** 448-454.

PAGANO, M., and K. GAUVREAU, 2000 Comparison of Two Means, pp. 259-284 in *Principles of Biostatistics*, edited by C. CROCKETT. Duxbury/Thompson Learning, Pacific Grove.

PANT, V., P. MARIANO, C. KANDURI, A. MATTSSON, V. LOBANENKOV *et al.*, 2003 The nucleotides responsible for the direct physical contact between the chromatin insulator protein CTCF and the *H19* imprinting control region manifest parent of

origin-specific long-distance insulation and methylation-free domains. Genes Dev **17:** 586-590.

PERK, J., K. MAKEDONSKI, L. LANDE, H. CEDAR, A. RAZIN *et al*., 2002 The imprinting mechanism of the Prader-Willi/Angelman regional control center. Embo J **21:** 5807-5814.

PREIS, J. I., M. DOWNES, N. A. OATES, J. E. RASKO and E. WHITELAW, 2003 Sensitive flow cytometric analysis reveals a novel type of parent-of-origin effect in the mouse genome. Curr Biol **13:** 955-959.

RAKYAN, V. K., S. CHONG, M. E. CHAMP, P. C. CUTHBERT, H. D. MORGAN *et al*., 2003 Transgenerational inheritance of epigenetic states at the murine *Axin*(Fu) allele occurs after maternal and paternal transmission. Proc Natl Acad Sci U S A **100:** 2538-2543.

REDDIEN, P. W., S. CAMERON and H. R. HORVITZ, 2001 Phagocytosis promotes programmed cell death in *C. elegans*. Nature **412:** 198-202.

REIK, W., A. COLLICK, M. L. NORRIS, S. C. BARTON and M. A. SURANI, 1987 Genomic imprinting determines methylation of parental alleles in transgenic mice. Nature **328:** 248-251.

REIK, W., and J. WALTER, 2001 Genomic imprinting: parental influence on the genome. Nat Rev Genet **2:** 21-32.

REINHART, B., M. ELJANNE and J. R. CHAILLET, 2002 Shared role for differentially methylated domains of imprinted genes. Mol Cell Biol **22:** 2089-2098.

REINKE, V., H. E. SMITH, J. NANCE, J. WANG, C. VAN DOREN *et al*., 2000 A global profile of germline gene expression in *C. elegans*. Mol Cell **6:** 605-616.

RUSSELL, D. W., and J. SAMBROOK (Editors), 2001 *Southern Hybridization of Radiolabeled Probes to Nucleic Acids Immobilized on Membranes*. CSHL Press, Cold Spring Harbor.

SAPIENZA, C., A. C. PETERSON, J. ROSSANT and R. BALLING, 1987 Degree of methylation of transgenes is dependent on gamete of origin. Nature **328:** 251-254.

SCHEDL, T., and J. KIMBLE, 1988 *fog-2*, a germ-line-specific sex determination gene required for hermaphrodite spermatogenesis in *Caenorhabditis elegans*. Genetics **119:** 43-61.

SCHEDL, T., 1997 Developmental genetics of the germline, pp. 241-269 in *C. elegans II*, edited by D. L. RIDDLE, T. BLUMENTHAL, B. J. MEYER and J. R. PREIS. Cold Spring Harbor Press, New York.

SCHNABEL, H., and R. SCHNABEL, 1990 An organ-specific differentiation gene, *pha-1*, from *Caenorhabditis elegans*. Science **250:** 686-688.

SCHOENHERR, C. J., J. M. LEVORSE and S. M. TILGHMAN, 2003 CTCF maintains differential methylation at the *Igf2*/*H19* locus. Nat Genet **33:** 66-69.

SCOTT, R. J., and M. SPIELMAN, 2004 Epigenetics: imprinting in plants and mammals--the same but different? Curr Biol **14:** R201-203.

SINGSON, A., 2001 Every sperm is sacred: fertilization in *Caenorhabditis elegans*. Dev Biol **230:** 101-109.

SLEUTELS, F., R. ZWART and D. P. BARLOW, 2002 The non-coding *Air* RNA is required for silencing autosomal imprinted genes. Nature **415:** 810-813.

SOPPE, W. J., S. E. JACOBSEN, C. ALONSO-BLANCO, J. P. JACKSON, T. KAKUTANI *et al.*, 2000 The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. Mol Cell **6:** 791-802.

STINCHCOMB, D. T., J. E. SHAW, S. H. CARR and D. HIRSH, 1985 Extrachromosomal DNA transformation of *Caenorhabditis elegans*. Mol Cell Biol **5:** 3484-3496.

STREISINGER, G., C. WALKER, N. DOWER, D. KNAUBER and F. SINGER, 1981 Production of clones of homozygous diploid zebra fish (*Brachydanio rerio*). Nature **291:** 293-296.

SWAIN, J. L., T. A. STEWART and P. LEDER, 1987 Parental legacy determines methylation and expression of an autosomal transgene: a molecular mechanism for parental imprinting. Cell **50:** 719-727.

THORVALDSEN, J. L., K. L. DURAN and M. S. BARTOLOMEI, 1998 Deletion of the *H19* differentially methylated domain results in loss of imprinted expression of *H19* and *Igf2*. Genes Dev **12:** 3693-3702.

TUCKER, K. L., C. BEARD, J. DAUSMANN, L. JACKSON-GRUSBY, P. W. LAIRD *et al.*, 1996 Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes. Genes Dev **10:** 1008-1020.

VERONA, R. I., M. R. MANN and M. S. BARTOLOMEI, 2003 Genomic imprinting: intricacies of epigenetic regulation in clusters. Annu Rev Cell Dev Biol **19:** 237-259.

VINKENOOG, R., C. BUSHELL, M. SPIELMAN, S. ADAMS, H. G. DICKINSON *et al.*, 2003 Genomic imprinting and endosperm development in flowering plants. Mol Biotechnol **25:** 149-184.

WALTER, J., and M. PAULSEN, 2003a The potential role of gene duplications in the evolution of imprinting mechanisms. Hum Mol Genet **12 Spec No 2:** R215-220.

WALTER, J., and M. PAULSEN, 2003b Imprinting and disease. Semin Cell Dev Biol **14:** 101-110.

WUTZ, A., O. W. SMRZKA, N. SCHWEIFER, K. SCHELLANDER, E. F. WAGNER *et al.*, 1997 Imprinted expression of the *Igf2r* gene depends on an intronic CpG island. Nature **389:** 745-749.

XIN, Z., C. D. ALLIS and J. WAGSTAFF, 2001 Parent-specific complementary patterns of histone H3 lysine 9 and H3 lysine 4 methylation at the Prader-Willi syndrome imprinting center. Am J Hum Genet **69:** 1389-1394.

YANG, Y., T. LI, T. H. VU, G. A. ULANER, J. F. HU *et al.*, 2003 The histone code regulating expression of the imprinted mouse *Igf2r* gene. Endocrinology **144:** 5658-5670.

YANOWITZ, J. L., M. A. SHAKIR, E. HEDGECOCK, H. HUTTER, A. Z. FIRE *et al.*, 2004 *UNC-39*, the *C. elegans* homolog of the human myotonic dystrophy-associated homeodomain protein Six5, regulates cell motility and differentiation. Dev Biol **272:** 389-402.

YOON, B. J., H. HERMAN, A. SIKORA, L. T. SMITH, C. PLASS *et al.*, 2002 Regulation of DNA methylation of *Rasgrf1*. Nat Genet **30:** 92-96.

CHAPTER 4

Probing the *C. elegans* genomic landscape using the *E. coli* DAM methylation system

Work on this project was initiated by JHU undergraduate Luiz Pantalena-Filho (who first cloned the *E. coli* dam methyltransferase into a *C. elegans* expression context), Carnegie Institution of Washington technician Jamie Fleenor (who inserted introns into the DAM coding region and injected the resulting constructs to produce a number of transgenic animals), JHU undergraduate Amy Goh (who carried out initial segregation and DNA preparations with strain PD5122), and Stanford technician Chaya Krishna (who carried out microinjections to produce transgenic line PD3994).

INTRODUCTION

DNA methylation is a recurrent theme in biology that is used by diverse organisms in various biological processes. In prokaryotes, C^5-methyl-cytosine (m5C), N^6-methyl-adenine (m6A), and N^4-methyl-cytosine (m4C) are common; whereas higher eukaryotes appear to possess mainly m5C systems (WION and CASADESUS 2006). All vertebrates and all angiosperms investigated thus far possess a DNA cytosine methylation system; while only certain species of fungi, invertebrates, protists, and prokaryotes possess a similar cytosine methylation system (GRACE GOLL and BESTOR 2005). In many plants and animals, DNA methylation is involved in the control of gene expression. For example, in mammals, cytosine methylation is important in the control of such processes as genomic imprinting and X chromosome inactivation (PRADHAN and ESTEVE 2003). Aberrant DNA methylation in mammals has been implicated in cancer (BAYLIN 2005). In plants, cytosine methylation is involved in the processes of genomic imprinting (FEIL and KHOSLA 1999; FINNEGAN *et al*. 1993), transcriptional silencing (KAPOOR *et al*. 2005; NEUHUBER *et al*. 1994), and paramutation (MEYER *et al*. 1993). Fungi likewise employ cytosine methylation as an epigenetic mark in repeat-induced point mutation (CAMBARERI *et al*. 1991).

In bacteria, m6A methylation is involved in a number of biological processes, including DNA replication, control of gene expression, mismatch repair, restriction-modification immunity (PALMER and MARINUS 1994), and bacterial virulence (HEITHOFF *et al*. 1999; OZA *et al*. 2005). *E. coli* also possess two stand-alone methyltransferases which are not part of any restriction-modification system. *dcm* (*D*NA *c*ytosine *m*ethylation) is a m5C methyltransferase with the target sequences CCAGG and CCTGG;

while *dam* (*D*NA *a*denine *m*ethylation) is a m6A methyltransferase with the target sequence GATC (MARINUS and MORRIS 1973; PALMER and MARINUS 1994).

Studies indicate that while vertebrates exhibit genome-wide methylation, invertebrate genomes are only sparsely methylated (TWEEDIE *et al.* 1997). This observation should be taken cautiously, however. Invertebrates far outnumber vertebrates both in the number of species and the diversity of biological processes. It is possible that, by coincidence, the limited number of invertebrate species examined thus far have very low DNA methylation activity or lack DNA methylation altogether. Insects appear to be the exception. Not only has cytosine methylation been detected in several insect species (FIELD *et al.* 2004), exotic DNA methylation such as 7-methylguanosine (in additional to N^6-methyl-adenosine) has been detected in mealybugs (ACHWAL *et al.* 1983; DEOBAGKAR *et al.* 1982). Additionally, the entire paternal genome of mealybugs become heterochromatic during embryogenesis (BONGIORNI and PRANTERA 2003; BROWN 1959; BROWN 1961; BROWN and NELSEN-REES 1961; BROWN and CHANDRA 1977; KHOSLA *et al.* 2006; SCHRADER and HUGHES-SCHRADER 1931). This heterochromatization has been shown to involve DNA methylation (BONGIORNI *et al.* 1999).

Despite the prevalence of DNA methylation in numerous species spanning diverse taxa, DNA methylation has not been detected in *C. elegans* (SIMPSON *et al.* 1986) and related nematode species. In fact, despite the great number of worm species spanning three phyla (Nematoda, Annelida, and Platyhelminthes), DNA methylation has been reported for only one annelid (segmented worm) species (DEL GAUDIO *et al.* 1997). Rarity of DNA methylation in worms is supported by phylogenetic analysis, which failed to uncover any evidence of a *dnmt-2*-like gene in *C. elegans* and the closely related

species *C. briggsae*, although a *dnmt-2*-like methyltransferase was detected in the free-living nematode *Pristionchus pacificus* (GUTIERREZ and SOMMER 2004). *dnmt-2* is one of several cytosine methyltransferases found in mammals and other eukaryotes that is highly conserved (VAN DEN WYNGAERT *et al.* 1998; YODER and BESTOR 1998). Gutierrez *et al.*, however, did find an *mbd*-like (*m*ethyl CpG *b*inding *d*omain) gene in all three nematode species they investigated. The highly conserved *mbd* domain is believed to be involved in recognition of methylated cytosine residues and is a component of DNA methylation systems in various species with cytosine methylation (WADE 2001). Loss-of-function of the *C. elegans mbd*-like gene resulted in a pleiotropic phenotype; whereas loss-of-function of the *C. briggsae mbd*-like gene resulted in paralyzed worms. Additionally, the *Caenorhabditis mbd*-like genes show less sequence conservation to other (highly conserved) *mbd*-like genes in other species that contain a DNA methylation system. Based upon these observations, Gutierrez *et al.* hypothesized that both *C. briggsae* and *C. elegans* may have lost their DNA methylation system and that components of the system may have evolved to perform other functions (GUTIERREZ and SOMMER 2004).

The *C. elegans* genome contains about100 megabases and is estimated to encode over 19,000 genes (CONSORTIUM 1998). The genome is packaged into five autosomes plus an X chromosome. The genome shows a high propensity for AA/TT dinucleotides repeated at regular, multiple intervals of 10 bases (FIRE *et al.* 2006). Such "phasing" of AA/TT dinucleotides is a unique feature of the *C. elegans* genome thus far not observed in non-nematode organisms. Multiple genetic studies have indicated that the landscape of *C. elegans* chromosome is not uniform. In particular, the autosomal centers have higher

gene densities (WATERSTON *et al.* 1992); while the autosomal arms have fewer genes, a higher recombination frequency than the centers, and are more repeat-rich (BARNES *et al.* 1995; CONSORTIUM 1998; GREENWALD *et al.* 1987; PRASAD and BAILLIE 1989; STARR *et al.* 1989).

We took advantage of the *E. coli* dam methyltransferase to construct an *in vivo* methylation system to probe features of the *C. elegans* genome. The methyltransferase is driven by the *myo-3* promoter. The *myo-3* gene codes for myosin heavy chain A, which is one of numerous structural proteins that comprise the *C. elegans* bodywall muscle system (DIBB *et al.* 1989; MILLER *et al.* 1986). *myo-3* promoter activity starts during mid-embryogenesis and continues through adulthood (HARFE *et al.* 1998; HONDA and EPSTEIN 1990; OKKEMA *et al.* 1993).

Using isoschizomeric restriction enzymes that are sensitive to the methylation status of the target sequence GATC, we show that the genomes of transgenic animals are methylated randomly throughout the entire genome. The methylation does not appear to be specific to muscle genes. Since dam methyltransferase is exogenous to *C. elegans*, it is unlikely to carry out a specific regulatory role in transgenic animals. We assume that target site methylation would reflect accessibility of the methyltransferase to chromatin. We had no a priori basis to predict either the degree of differences in accessibility or the extent to which dam methylation might affect development or physiology of transgenic animals. Hence, the artificial methylation system we have constructed would be a useful tool in studying *C. elegans* chromatin *in vivo* without affecting development of the animal.

MATERIALS AND METHODS

C. elegans strains and growth conditions

Animals were reared according to standard protocols on *E. coli* grown on NGM (nematode growth medium) nutrient plates (BRENNER 1974). N2 (wildtype) animals were fed *E. coli* strain OP50 and/or SCS110. PD5122 animals were fed *E. coli* strain SCS110. PD3994 animals were fed *E. coli* strain SCS110(Amp^R), which is an ampicillin-resistant derivative of SCS110 made by transformation with pUC18. *E. coli* strain SCS110 (*rpsL* (Strr) *thr leu endA thi-1 lacY galK galT ara tonA tsx dam dcm supE44D* (*lac-proAB*) [F' *traD36 proAB lacIqZDM15*]) is defective in both *dcm* and *dam* methylation; while *E. coli* strain OP50 is wildtype for both methylation systems. Both SCS110 and SCS110(Amp^R) were periodically tested on 2-amino-purine (2AP) to ensure that the stock has not been contaminated with dam$^+$ bacteria. *E. coli* that are dam$^-$ grow poorly in the presence of 2AP compared to dam$^+$ *E. coli* (PALMER and MARINUS 1994). All worm strains were reared at 23°C unless otherwise specifically stated. *C. elegans* strains used in the experiments were as follows:

N2: wildtype strain of *C. elegans* (Bristol isolate) (BRENNER 1974)

PD3994 [*pha-1*(*e2123ts*) III; *ccEx3994*]: transgenic line expressing the *E. coli* dam methyltransferase and genomic *C. elegans pha-1*(+) gene (SCHNABEL and SCHNABEL 1990) from the extra-chromosomal array *ccEx3994*[pPD176.59 + pC1].

PD5122 [*pha-1*(*e2123ts*) III; *ccEx5122*]: transgenic line expressing the *E. coli* dam methyltransferase-GFP translational fusion and genomic *C. elegans pha-1*(+) gene from the extra-chromosomal array *ccEx5122*[pPD177.01 + pC1]. PD5122 worms show strong and uniform GFP expression in the bodywall muscles.

Plasmids and transgenic lines

Line PD5122 was established by microinjection of a mixture of plasmids pPD177.01 (Lig6682) and pC1 into *pha-1*(*e2123ts*) animals (GRANATO *et al*. 1994). pPD177.01 contains the *myo-3* (bodywall muscle) promoter driving the *E. coli* dam methyltransferase fused to GFP. Two introns with *C. elegans* consensus sequences have been inserted into the dam methyltransferase to optimize expression in nematodes (J. Fleenor and A. Fire, personal communication). pC1 contains the *C. elegans* genomic *pha-1* gene and is a selection marker for the transgene. Line PD3994 was established by microinjection of plasmids pPD176.59 (Lig6649) and pC1. pPD176.59 contains the *E. coli* dam methyltransferase driven by the *myo-3* promoter and a single SV40 nuclear localization signal. Plasmid pPD98.38 (Lig2524) carries a single copy of the 5S/SL1 rDNA sequence from which the Southern blot probe was synthesized.

Southern blots

Southern blots were performed according to standard protocols. Briefly, RNase A-treated genomic DNA were subjected to a one hour restriction digest by *Dpn I*, *Mbo I*, or *Sau3A I*, followed by phenol:chloroform extraction and ethanol precipitation. Restricted fragments were separated on 1.4% agarose gels followed by transfer to Hybond-N+ membranes (Amersham Biosciences, Cat. #RPN303B). We used the method of capillary blotting under alkali conditions. Radio-labeled probes 808bp long containing the *C. elegans* 5S rDNA/SL1 (Spliced Leader Sequence 1) were synthesized from a *Bam* H1 fragment of plasmid pPD98.38 using the RadPrime DNA Labeling System (Invitrogen, Cat. #18428-11) with radio-labeled α-^{32}P dATP (MP Biomedicals, Cat. #:33002HD.5). Southern hybridization was performed as described in Molecular

Cloning: A Laboratory Manual (RUSSELL and SAMBROOK 2001). We performed pre-hybridization and hybridization in roller bottles using phosphate-SDS buffer [0.5M phosphate buffer (pH7.2), 1mM EDTA (pH8.0), 7% (w/v) SDS, 1% (w/v) BSA]. Membranes were exposed to Kodak BioMax XAR Film (Cat. 165-1454) for the appropriate amount of time.

Cloning and Sequencing of *Dpn I* genomic fragments

RNase A-treated genomic DNA was digested for one hour by *Dpn I* in a 200µL reaction volume. Digestion was stopped by the addition of 350µL STOP buffer (1M NH4Ac, 10mM EDTA,100ug/mL tRNA,0.2% SDS) followed by phenol:chloroform extraction (500µL), chloroform extraction (500µL), and ethanol precipitation (1mL 100%), an ethanol wash (100%), and resuspension in pH 7.4 TE. Digested products were ligated into the pZErO-2 vector (Cat. K2600-01, Invitrogen) at the *EcoR V* site (blunt-ended), transformed into TOP10 competent cells (Invitrogen, #44-0301) and grown overnight at 37°C. Each colony was picked and grown in culture overnight, mini-prepped, and digested with *Pvu II* to indicate the size of the insert before being sent for sequencing at Elim Biopharmaceuticals (Hayward, California). We sequenced primarily from the T7 priming site using T7 forward primers on the pZErO-2 vector. For longer sequences, we also sequenced from the M13 priming site using M13 reverse primers. Sequences were aligned against the *C. elegans* genome (March 2004 assembly) using BLAT software (HINRICHS *et al.* 2006; KENT 2002) from UC Santa Cruz (http://genome.ucsc.edu/cgi-bin/hgBlat). From the BLAT alignment, we pared off flanking vector sequences and used the resulting sequence to determine the corresponding gene and physical map position of the fragment.

SAGE analysis

SAGE data were obtained from the Genome BC *C. elegans* Gene Expression Consortium http://elegans.bcgsc.bc.ca/ (McKay *et al*. 2003). We downloaded the March 2006 *C. elegans* SAGE database using the following relatively standard parameters: Quality filter: 0.99, Hide ambiguous tags: ON, Tag mapping resources: CODING, Show only mapped tags:ON, Tags/page: 10, Lowest count cutoff: 1, Hide antisense tags: ON, Remove duplicate ditags: ON, Highest count cutoff: NONE, Sort order: DOWN, Resolve lowest match: ON.

RESULTS

To gain deeper insight into *C. elegans* chromatin, we asked which regions of the genome are susceptible to DNA modification, the rationale being that more compacted regions of the genome (i.e. heterochromatin) may be less accessible to chromatin modifying factors (Figure 4.1). To this end, we constructed an experimental system by taking advantage of the *E. coli* dam methyltransferase. Dam methyltransferase transfers a methyl group from S-adenosyl-methionine to adenine in the target sequence GATC. By mapping sites of DAM methylation, we should be able to determine relatively open and condensed chromatin regions of the *C. elegans* genome.

Engineering *E. coli* dam methyltransferase for expression in *C. elegans*

We've previously found that introns incorporated into the coding region can improve expression of transgenes in *C. elegans*. The dam methyltransferase gene was cloned by PCR from *E. coli* strain OP50, and two introns were incorporated at blunt cutting restriction sites. The resulting coding region was placed behind the *myo-3* promoter, which expressed in *C. elegans* bodywall muscles (ARDIZZI and EPSTEIN 1987; OKKEMA *et al.* 1993). Two constructs were produced. In one, pPD177.01, the dam methyltransferase was designed to express as a fusion to GFP; while in the other, pPD176.59, a nuclear localization signal from SV40 was included. Each construct was then incorporated into transgenic strains using *pha-1*(+) as a selectable genetic marker in a *pha-1*(*e2123ts*) genetic background.

Transgenic animals do not show any perceptible morphological defects

In the two transgenic lines we have constructed, dam methyltransferase activity in *C. elegans* muscle appears to be compatible with cellular function and growth. These

transgenic animals look normal morphologically. In a blind test, the transgenic animals move slightly slower than wildtype. The growth rate is slower than wildtype animals. However, we do not know if the slow growth is due to the *pha-1*(*e2123ts*) background and/or expression of dam methyltransferase.

Detecting expression of dam::GFP fusion protein by fluorescence microscopy

Line PD5122, which expresses the dam::GFP fusion protein, was examined under fluorescence microscopy to determine the tissue-specificity and distribution of dam methyltransferase (Figure 4.2). When the animal is oriented laterally as in panels A (adult) and D (L4 larvae), the bodywall muscles run along the animal's body dorsally and ventrally. Notice that the central dark regions, comprising the gut and gonad arms, are devoid of GFP. Within the body musculature, there is some degree of mosaicism in transgene expression in these animals. For example, notice in Panel A that certain bodywall muscle cells (arrow) lack GFP, possibly due to loss of the extra-chromosomal transgene array. Panels E and F show the bodywall muscles from dorsal and ventral views, respectively. In addition to bodywall muscles, *myo-3* is also expressed in the vulval muscles (ARDIZZI and EPSTEIN 1987; OKKEMA *et al.* 1993), easily identified as an "X" on the animal's ventral surface (panels F and G). Panels B, C, and H are close-ups of bodywall muscles from various regions of the animal body. Notice the presence of small granules in all these muscle cells. We do not know the identity of these punctate structures, but we do not think they are mitochondria, as there are far more mitochondria in *C. elegans* bodywall muscle cells than what is seen in these pictures (FIRE *et al.* 1998). Since the dam::GFP fusion used in PD5122 does not contain a characterized nuclear localization signal, we were initially unsure whether the fusion protein would have access

to the nucleus. However, as Figure 4.2 shows, not only does the methyltransferase have access to the nucleus, it appears to be found predominantly in the nucleus. Also note in Panels B, C, and H the absence of a "dark spot" within the nucleus, possibly indicating that the dam methyltransferase may be localized in the nucleolus as well.

DAM can methylate DNA at a high fraction of potential target sites

Using three restriction enzymes that target GATC in a methylation-dependent manner (Figure 4.3), we show that there is a clear difference in methylation state between genomes of animals expressing dam methyltransferase versus the wildtype control. All three enzymes have the same target sequence: *Dpn I* cuts only methylated GATC; *Mbo I* cuts only non-methylated GATC; while *Sau3A I* cuts both methylated and non-methylated GATC (Figure 4.3).

Two transgenic lines were available for analysis. PD3994 expresses a nuclear-localized dam methyltransferase; while PD5122 expresses a dam::GFP translational fusion that does not contain a nuclear localization signal. We digested wildtype and transgenic genomes using *Dpn I*, *Mbo I*, or *Sau3A I*, and probed with an ≈800bp probe with sequences from the *C. elegans* 5S/SL1 rDNA locus. The *C. elegans* 5S/SL1 rDNA locus consists of about 110 copies of 1kb repeats and generally has high transcriptional activity (KRAUSE and HIRSH 1987; NELSON and HONDA 1989).

Figures 4.4 and 4.5 show that the methylation state of the two genomes expressing dam methyltransferase differ from that of the N2 (wildtype) control. In Figure 4.4A, lane PD3994 is clearly different from the N2 lane. The former shows a strong smear that is nearly absent in N2, indicating that methylated DNA substrates were available for the *Dpn I* enzyme. This difference is even more evident in the

corresponding Southern blot probed with Ce5S/SL1 (Figure 4.4B). Distinct bands are present in the PD3994 lane that are clearly absent in the N2 lane. These same patterns are also seen for PD5122 in Figures 4.5A and 4.5B.

The slight smearing seen in the ethidium bromide gels for the N2 lanes in Figures 4.4A and 4.5A likely results from the non-specific activity of *Dpn I*, which can cut non-methylated GATC 1,000-fold more slowly than fully methylated GATC (and hemi-methylated DNA at a 60-fold lower rate). [New England Biolabs, personal communication]

Restriction by *Mbo I* likewise revealed differences between the DAM-exposed and wildtype genomes. In Figures 4.4C and 4.5C, we see that *more* of the genomic DNA in PD3994 (Figure 4.4C, arrow) and PD5122 (Figure 4.5C, arrow) are left uncut by *Mbo I* compared to N2 DNA. That there is smearing in the agarose gels for PD3994 and PD5122 in the *Mbo I* digests can be explained by the fact that both *ccEx3994* and *ccEx5122* are driven by a muscle promoter and expressed specifically in muscle; thus, in these transgenic animals, the non-methylated DNA in non-muscle tissues (comprising about 90% of the body mass) are substrates for *Mbo I*. In the corresponding Southern blots for *Mbo I* digests (Figures 4.4D and 4.5D), we see that N2 DNA is essentially completely digested whereas PD3994 and PD5122 DNA are only partially digested. The unrestricted bands (indicated by arrows) in Figures 4.4D and 4.5D are likely methylated DNA from muscle tissue.

Since *Sau3A I* cuts both methylated and non-methylated GATC, we expect that *Sau3A I* restriction would produce identical banding patterns between wildtype (non-

methylated) DNA and PD3994/PD5122 (methylated) DNA. This is indeed the case, as shown in Figures 4.4D and 4.5F.

DAM expression leads to modifications throughout the genome

To map DAM-targeted sequences in the two transgenic lines, we cloned *Dpn I* digested genomic fragments from the two lines according to the procedure described in Materials and Methods. The results of the cloning experiments are summarized in Table 4.1. The cloning experiments yield a total of 182 and 174 sequences fragments for PD3994 and PD5122, respectively. These sequences span exons, introns, exon/intron junctions, and non-annotated (intergenic) regions. After discarding redundant clones, we obtained 168 unique sequences for PD3994 and 167 unique sequences for PD5122. We do not believe the redundant clones to be independent cloned fragments. Rather, they likely resulted from *E. coli* duplication during the growth phase in culture. Of the unique sequences, we were able to confirm 153 from PD3994 and 161 from PD5122 to be *Dpn I* fragments. We categorize a sequence (after flanking vector sequences were pared off) as potentially resulting from a *Dpn I* restriction if it is flanked by TC on one side and GA on the other side. Clone #115 from PD3994 carried a non-*Dpn I* fragment that corresponded to a 21bp sequence from the pZErO-2 vector. Unexpectedly, we also cloned mitochondrial sequences from both PD3994 and PD5122. Seven sequences, corresponding to five unique mitochondrial genes, were cloned. Six of these mitochondrial sequences are confirmed *Dpn I* fragments, and one has a single internal GATC site.

We performed an N2 control in tandem with the PD3994 cloning experiment. DNA from *C. elegans* strain N2 (lacking the dam methyltransferase) grown on

dam-minus bacterial strain SCS110(Amp^R) was digested with *Dpn I* and cloned as above. We noted that many fewer colonies (at least 10-fold fewer) were recovered from the N2 sample. This is consistent with the expected inability of *Dpn I* to cut N2 DNA. Of the small number of N2 clones obtained, twenty-four were digested with *Pvu II* to check the size of the inserts. Twenty-three of 24 N2 clones gave patterns indicating the absence of any insert or inserts that were too short to be detected on the agarose gel (data not shown). We sequenced N2 clones 1-5. Clones 1, 3, and 5 had the identical (and presumably artifactual) 21bp inserts from the pZErO-2 vector (PD3994 clone #115 had a similar 21bp insert). N2 clone #4 had a 33bp insert corresponding to a sequence from Chromosome II. And N2 clone #2 had a 312bp insert corresponding to a sequence on Chromosome X. None of the N2 clones were *Dpn I* fragments.

Tables 4.2 and 4.3 summarize the annotated sequences that correspond to cloned *Dpn I* fragments. A substantial fraction of *Dpn I* fragments from both lines contained internal GATC sites. In general, the number of internal GATC sites is proportional to the length of the fragment (Figure 4.6). As is evident in Figure 4.7, *Dpn I* fragment sizes varied widely, from 25bp to over 6kb. In almost all cases, we were able to corroborate the listed fragment size using a *Pvu II* digest of the plasmid mini-prep (data not shown). Figure 4.7 shows the distribution of *Dpn I* fragment sizes from the cloning experiments. Although the range of sizes is quite large, the majority of fragments are 100-1,000bp long.

Figure 4.8 shows the genomic distribution of *Dpn I* fragments from the two cloning experiments from DAM-expressing animals. The figure shows that our cloning

experiments were not biased towards particular regions of the genome. Thus, we sampled a wide representation across each chromosome and across the entire *C. elegans* genome.

Methylation is not limited to muscle genes

Since the dam methyltransferase in both lines were expressed under the control of a muscle promoter, the distribution of accessible fragments also allows us to compare at least one aspect of expression with genome accessibility. It was certainly conceivable (and would have been of interest) if only expressed regions of the genome were accessible to dam methyltransferase. Alternatively, expression might have little relationship with DAM accessibility.

For each *Dpn I* fragment, we assigned a score to the nearest gene based upon its proximity (Figure 4.9). We then determined that gene's weighted "hit frequency" for each characterized tissue in the *C. elegans* SAGE database (McKay *et al*. 2003). Finally, we determined the average weighted SAGE score for each tissue or all genes containing cloned *Dpn I* fragments, and compared these values to values for all genes in the SAGE dataset.

The results of the SAGE analysis are shown in Tables 4.4 and 4.5. Targets of the dam methyltransferase do not appear to be specific to muscle genes. In fact, the SAGE data indicate a global target of the methyltransferase, as the majority of SAGE hits for each tissue from the cloning experiments do not appear to show large statistical deviations from the wildtype control. These results point to a situation in which, within muscle cells, dam methyltransferase has access to the entire genome.

DISCUSSION

In this chapter, we have described the development of a method to take advantage of the *E. coli* dam methylation system to probe features of *C. elegans* chromatin. *E. coli* dam methyltransferase adds a methyl functional group to adenine in the target sequence GATC. In conjunction with the dcm methylation system (which methylates cytosine residues), adenine methylation in *E. coli* is used in a number of biological processes, including virulence (EROVA *et al*. 2006; FALKER *et al*. 2005; HEITHOFF *et al*. 1999; OZA *et al*. 2005), immunity against phages, and regulation of various cellular activities (BARRAS and MARINUS 1989; PALMER and MARINUS 1994; WION and CASADESUS 2006).

To gain insight into *C. elegans* chromatin organization, we engineered two transgenic lines expressing *E. coli* dam methyltransferase from extra-chromosomal transgenes. In both lines, the methyltransferase is under the control of the *myo-3* promoter, which codes for a structural component of the bodywall musculature. In line PD3994, the methyltransferase is fused to a characterized nuclear-localization signal (from SV40); whereas in line PD5122, the methyltransferase is simply fused to GFP. Unexpectedly, the DAM-GFP fusion protein in PD5122 localizes to the nucleus with high efficiency (Figure 4.2B,C,H), even though it is not predicted to be nuclear-localized.

One advantage of this experimental system is that DNA methylation does not normally exist in *C. elegans*. An exogenous system would likely not be regulated with regard to targeting of the methyltransferase to particular regions of the genome. Thus, regions of the genome targeted by the methyltransferase would presumably be due to chromatin organization. Another advantage of our experimental system is in allowing us to study chromatin in a physiological context.

Our analysis of transgenic animals expressing dam methyltransferase shows that there is methylation of *C. elegans* DNA in the corresponding genomes. Figures 4.4 and 4.5 are Southern blots that show the methylation status at the 5S rDNA locus is clearly different between N2 and the DAM-expressing lines. That we were able to easily clone *Dpn I* fragments from the two DAM-expressing lines but not from N2 was clear evidence that each of the transgenic lines had functional DNA methyltransferase activity.

In *C. elegans*, there are 770 potential targets of dam methyltransferase at the 5S rDNA cluster. The Southern blots show that restriction at this cluster produces virtually identical patterns between PD3994 and PD5122. This is true for both *Dpn I* and *Mbo I* restriction. At present we do not have an explanation for this observation. But it does suggest a reproducible pattern between independent lines expressing dam methyltranferase, owing perhaps, to accessibility of the methyltransferase to the same target sites.

The targets of the dam methyltransferase are not limited to DNA regions expressed in muscle tissue (SAGE data in Tables 4.4 and 4.5). This result is somewhat unexpected, as the methyltransferase is driven from a muscle promoter. One explanation for the lack of tissue specificity could be that the methyltransferase was active during the initial phase of genomic DNA extraction, when worms were lysed. This would allow the methyltransferase to methylate the DNA of non-muscle cells. However, we do not think this to be the case. First, worm lysis was performed at 62°C in the presence of EDTA, conditions that would have been extremely hostile to a protein whose optimal activity is at 37°C (dam methyltransferase is inactivated at 65°C). In our observations of dam::GFP fusion expression in PD5122, the GFP is localized exclusively to muscle cells (Figure

4.2). Thus, it is likely that dam methyltransferase has access to much of the genome in muscle, and any chromatin accessibility differences between muscle and non-muscle genes are relatively subtle. This implies that dam methyltransferase targets genes in muscle tissue independent of their transcriptional activities, and that expression of dam methyltransferase from a single tissue type can be used to analyze global gene expression. GATC methylation in the two transgenic lines does indeed appear to be global, as we were able to clone *Dpn I* fragments from disparate regions of the genome (Figure 4.8).

We note that a fraction of the cloned *Dpn I* fragments contains one or more internal GATC sites. In general, the number of internal GATC sites is proportional to the length of the fragment (Figure 4.6). *E. coli* dam methyltransferase has been shown to be a highly processive enzyme. It functions as a monomer and methylates DNA one strand at a time (URIG *et al*. 2002). DAM is able to methylate about 55 target sites in a single binding event, as assayed *in vitro* on λ DNA (URIG *et al*. 2002). Hence, the result of dam methyltransferase activity is such that a group of consecutive targets is either hemi-/fully methylated or it is left unmethylated on both strands. If this mechanism of DAM operated in our transgenic animals, then the presence of internal GATC sites in certain *Dpn I* fragments could be due to exclusion of dam methyltransferase access by higher order chromatin structures. Alternatively, the presence of internal DAM targets in certain *Dpn I* fragments could be due to stochasticity in *Dpn I* restriction. Because *Dpn I* has low activity for non-methylated GATC (three orders of magnitude less compared to fully methylated GATC) and with limited 1/60-fold activity on hemi-methylated DNA, we limited the *Dpn I* genomic DNA digests to one hour. Although this would be predicted to

(and apparently did) give us near complete digestion of fully methylated DNA, we do not know to what extent hemi-methylated DNA would be cut in such assays. Hence, the internal DAM targets sites in some *Dpn I* fragments could be hemi-methylated GATCs, for which *Dpn I* could not efficiently cut. Thus, our cloning strategy does not distinguish between hemi-methylated and non-methylated GATC.

The nuclear genome of *C. elegans* contains over 269,000 GATC sites, with potential *Dpn I* fragments ranging from a few bases to over 31,000 bases long, and an average *Dpn I* fragment length of 369bp. Our two manual cloning experiments sampled only 308 (0.11%) potential targets of dam methyltransferase. The small sample size does not have the required resolution to allow us to draw a conclusion on whether reproducible chromatin structure and/or stochasticity plays a role in excluding DAM from certain target sites.

There is precedent for restriction of DNA methylation to certain genomic landscapes by chromatin structure. In *Tetrahymena*, chromatin structure appears to play an essential role in determining the endogenous DNA methylation status of potential target sites. For example, endogenous adenine methylation at GATC sites is dependent upon nucleosomal positioning, where methylation preferentially occurs at linker DNA (KARRER and VANNULAND 2002). Also, GATC methylation in *Tetrahymena* appears to have a positional effect. Previously methylation competent sites, when translocated into certain genomic regions, become methylation incompetent (KARRER and VANNULAND 1998). Thus, in *Tetrahymena*, chromatin structure appears to be a stronger determinant of methylation competency than target sequence. Although local chromatin structure seems

to determine methylation status in *Tetrahymena*, there is no evidence for inaccessibility to methylation in transcriptionally inactive regions (KARRER and STEIN-GAVENS 1990).

The *C. elegans* genomic landscape contains nucleosome::DNA interactions that may limit the access of dam methyltransferase to certain GATC sites. Johnson *et al.* (submitted) have recently shown that nucleosome cores in bulk *C. elegans* chromatin, although they may be highly positioned in a limited set of cases, show considerable heterogeneity in their positions. There may be regions of the genome with greater constraint, as suggested by bioinformatic analysis (FIRE *et al.* 2006). Analysis of the susceptibility of these regions to dam methyltransferase will be of considerable interest.

We note that there is little difference between the nuclear-localized (Line PD3994) and non-nuclear-localized (line PD5122) dam methyltransferase. The methyltransferase in both lines appear to have equal access to the nuclear genome. For example, we did not encounter any difficulty cloning *Dpn I* fragments from PD5122, as would be the case if the non-nuclear localized methyltransferase were prevented from access to the nucleus. Also, the methylation profiles (distribution of *Dpn I* fragments in the genome and among chromosomes, distribution of *Dpn I* fragment lengths, and ability of DAM to target both muscle and non-muscle genes) are indistinguishable between the two transgenic lines. This suggests that, although methyltransferase activity must occur in the nucleus, a nuclear localization signal is not required for the methyltransferase to reach its target in the nucleus. In wildtype *E. coli*, only about 130 molecules of dam methyltransferase are present in each cell (BOYE *et al.* 1992). This number of molecules is apparently sufficient to carry out the full complement of *E. coli* cellular activities that require dam methylation. Hence, even if only a small number of the non-

nuclear-localized version of the methyltransferase manage to find their way into the nucleus, their activity may be very potent. In our observation of dam::GFP distribution within muscle cells, we have found a high level of the fusion protein present in the nucleus, despite the lack of a nuclear localization signal in the fusion protein. Thus, if only a small number of methyltransferase molecules is required to target all the potential GATC sites in the *C. elegans* genome, then the methyltransferase in our system may be saturated. And if this is the case, then the presence of internal GATC sites in our cloning experiments would implicate that dam methyltransferase was blocked from access to these sites.

Interestingly, we also cloned six unique mitochondrial *Dpn I* fragments from both transgenic lines. All six clones have very low sequence homology to sequences in the nuclear genome and they are not *E. coli* sequences. Hence, they are unequivocally of mitochondrial origin. There are certainly examples of cytosine (but not adenine) methylation of mtDNA in animals (KUDRIASHOVA *et al.* 1976). And in plants, N^6-adenine methylation of mtDNA was recently discovered (FEDOREYEVA and VANYUSHIN 2002). The enzyme, *wadmtase*, has the target sequence TGATCA. As neither cytosine nor adenine methylation activity has been detected in the nuclear genome of *C. elegans*, it is unlikely that *C. elegans* possesses endogenous mtDNA methylation (otherwise, we would have cloned mitochondrial *Dpn I* fragments from our N2 control). It is conceivable that there were traces of mtDNA present in the cytoplasm (possibly from fracture of mitochondria during the DNA extraction procedure) that were targeted by dam methyltransferase. Because *C. elegans* muscle cells contain very high mitochondria count (FIRE *et al.* 1998), this is a plausible explanation. The other alternative, that dam

methyltransferase can enter in the mitochondria, is unlikely for three reasons. First, proteins localized to the mitochondria require a mitochondrial localization signal. Second, although the outer mitochondrial membrane is somewhat permeable, the inner membrane is highly non-permeable and requires regulated entry or exit. Third, if dam methyltransferase somehow found its way into the mitochondria, we would have cloned more mitochondrial sequences.

We have not noticed an *uncoordinated* phenotype (indicative of defects in movement due to muscle dysfunction) in the two transgenic lines. The *myo-3* gene has a relatively strong promoter whose activity begins at mid-embryogenesis and continues through adulthood. Defects in myosin or in muscle function result in an *unc* phenotype for *C. elegans*. Since DNA demethylase activity has not been detected in any organism, we would expect that accumulation of methylated DNA in muscle cells might lead to defects in muscle function. It may require accumulation of methylated DNA in muscle tissue over multiple generations for any defect to become noticeable. Since the methyltransferase is driven by a somatic gene promoter, our transgenic animals would escape any such accumulated defects. We should be able to test this hypothesis with a transgenic line expressing dam methyltransferase driven by a germline gene promoter.

The lack of any perceptible defects in movement of the DAM-expressing animals could indicate that these transgenic animals are "blind" to DNA methylation, since *C. elegans* does not possess an endogenous DNA methylation system. Similar experiments in *Drosophila* support our view. Ectopic expression of dam methyltransferase in *Drosophila* did not cause any abnormalities in transgenic animals (WINES *et al.* 1996). However, ectopic expression of the mammalian 5-methyl-cytosine

methyltransferases *Dnmt3* (a *de novo* methyltransferase) and *Dnmt1* (a maintenance methyltransferase) resulted in severe developmental abnormalities, including lethality (LYKO *et al*. 1999). Because *Drosophila* possesses endogenous *dDnmt2* activity (a close homolog of mammalian *Dnmt2* 5-methyl-cytosine methyltransferase), it is likely that ectopic CpG methylation by *Dnmt3* and *Dnmt1* interacted with the endogenous methylation system.

In conclusion, we have constructed an experimental system to study *C. elegans* chromatin. Expression of the *E. coli* dam methyltransferase in *C. elegans* muscle leads to genome-wide N^6-adenine methylation. Transgenic animals do not appear to suffer from any noticeable phenotypes. Dam methyltransferase does not appear to hit targets in a tissue-specific manner, allowing DAM access to the entire genome from within muscle tissue. Combined with the results from the Tetrahymena work, our results in *C. elegans* suggest that analysis of DAM accessibility may serve as a valuable and relatively general tool to investigate numerous genomic sites for positioning relative to nucleosome cores.

FIGURES

Figure 4.1. Experimental logic behind use of *E. coli* dam methyltranferase to probe chromatin. When present in heterochromatin or bound by a protein(s), GATC sites (green) are inaccessible to dam methyltransferase. However, GATC sites located in euchromatin are accessible to the methyltransferase.

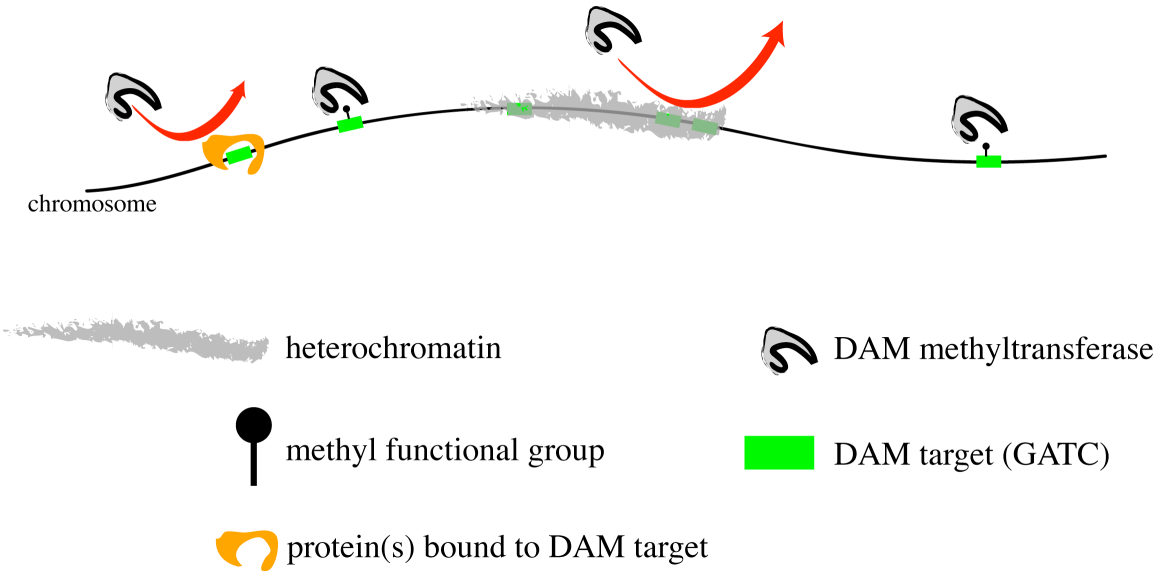


Figure 4.2. Tissue specific expression and distribution of dam methyltransferase in *C. elegans* muscle (line PD5122). (A, D) Lateral views of dam::GFP expression in the bodywall muscles of an adult and L4 larvae, respectively. (E, F) dorsal and ventral views, respectively, of a young adult. (F, G) *myo-3* is normally expressed in the vulval muscles ("X") as well. (B, C, H) Close-ups of various bodywall muscle cells. Animals were alive at time of observation. Pictures were taken using a compound microscope.

Figure 4.2

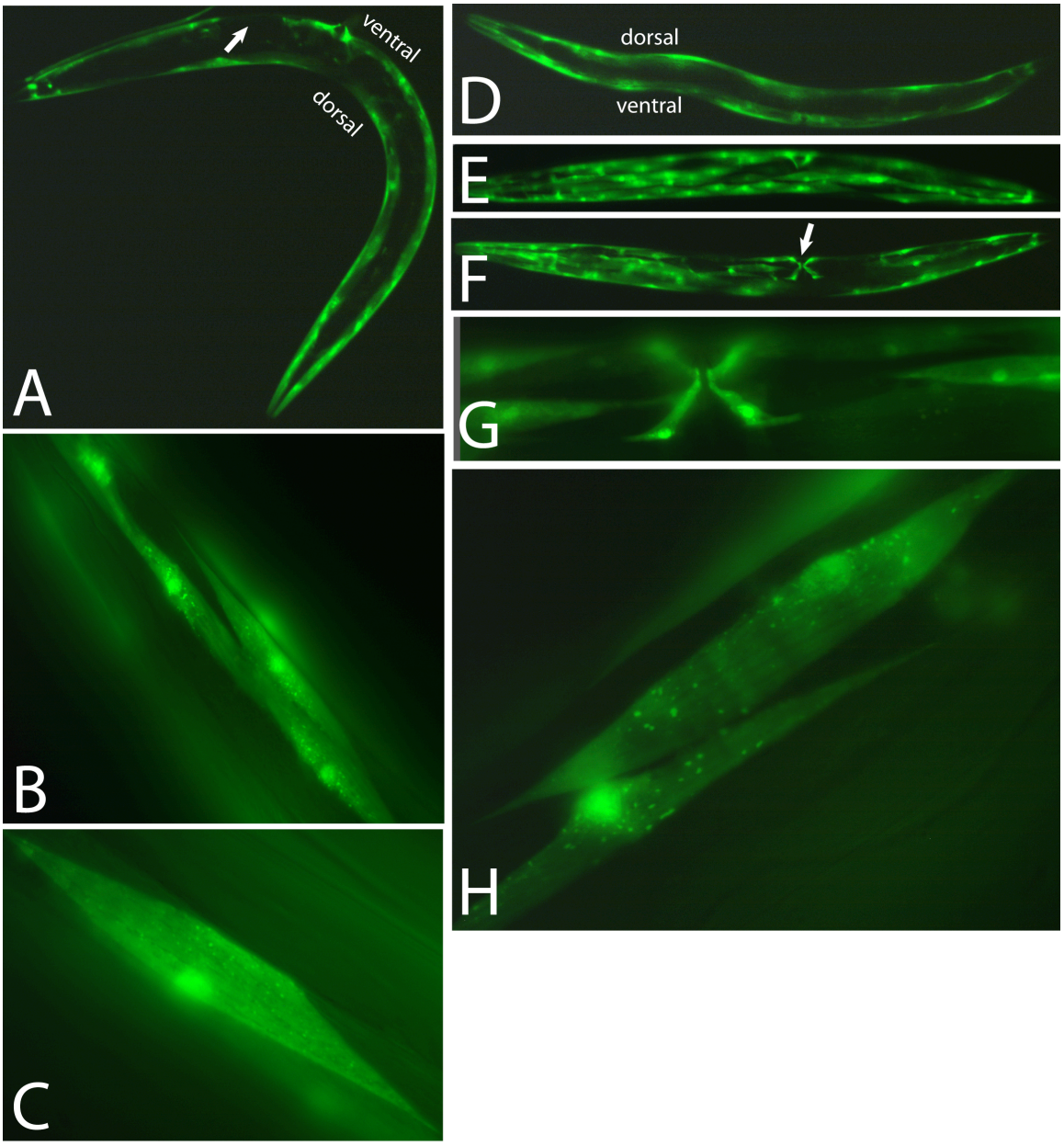


Figure 4.3. Specificity of the isoschizomers *Dpn I*, *Mbo I*, and *Sau3A I*. All three enzymes target the same sequence but the activities of *Dpn I* and *Mbo I* are dependent upon the methylation status of the target sequence. *Dpn I* generates blunt ends while *Mbo I* and *Sau3A I* generate sticky ends.

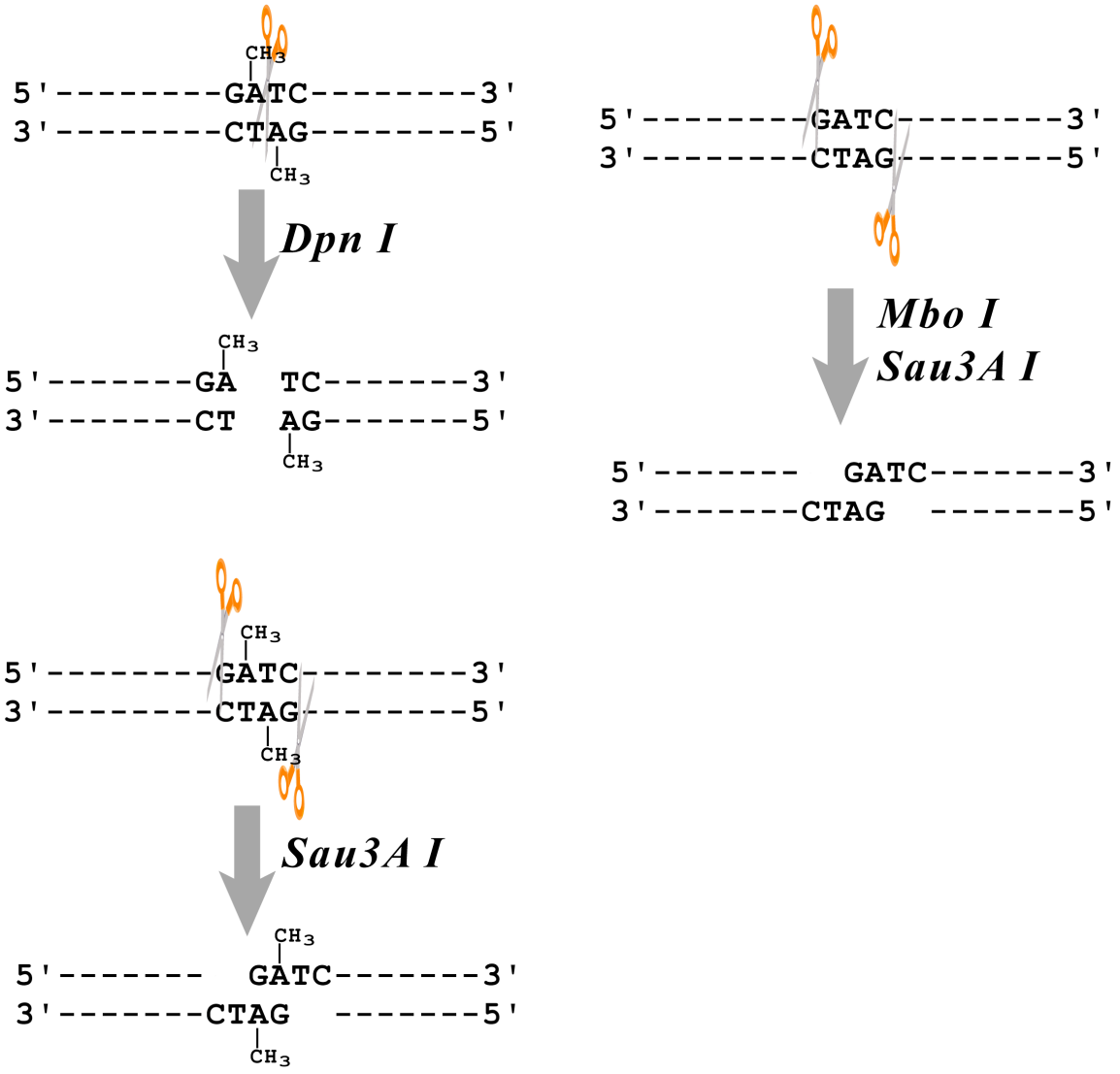


Figure 4.4. Southern blot analysis of PD3994 genomic DNA digested with *Dpn I*, *Mbo I*, or *Sau3A I*. (A, C) Ethidium bromide stained agarose gels. (B, D) The corresponding Southern blots to the agarose gels. Plasmid pPD98.38 was propagated in *E. coli* wildtype for dam methyltransferase and was, therefore, fully methylated. Hence, pPD98.38 was a substrated for both *Dpn I* and *Sau3A I*, but not *Mbo I*. The probe was a 808bp oligonucleotide containing the *C. elegans* 5S rDNA/SL1 sequence.

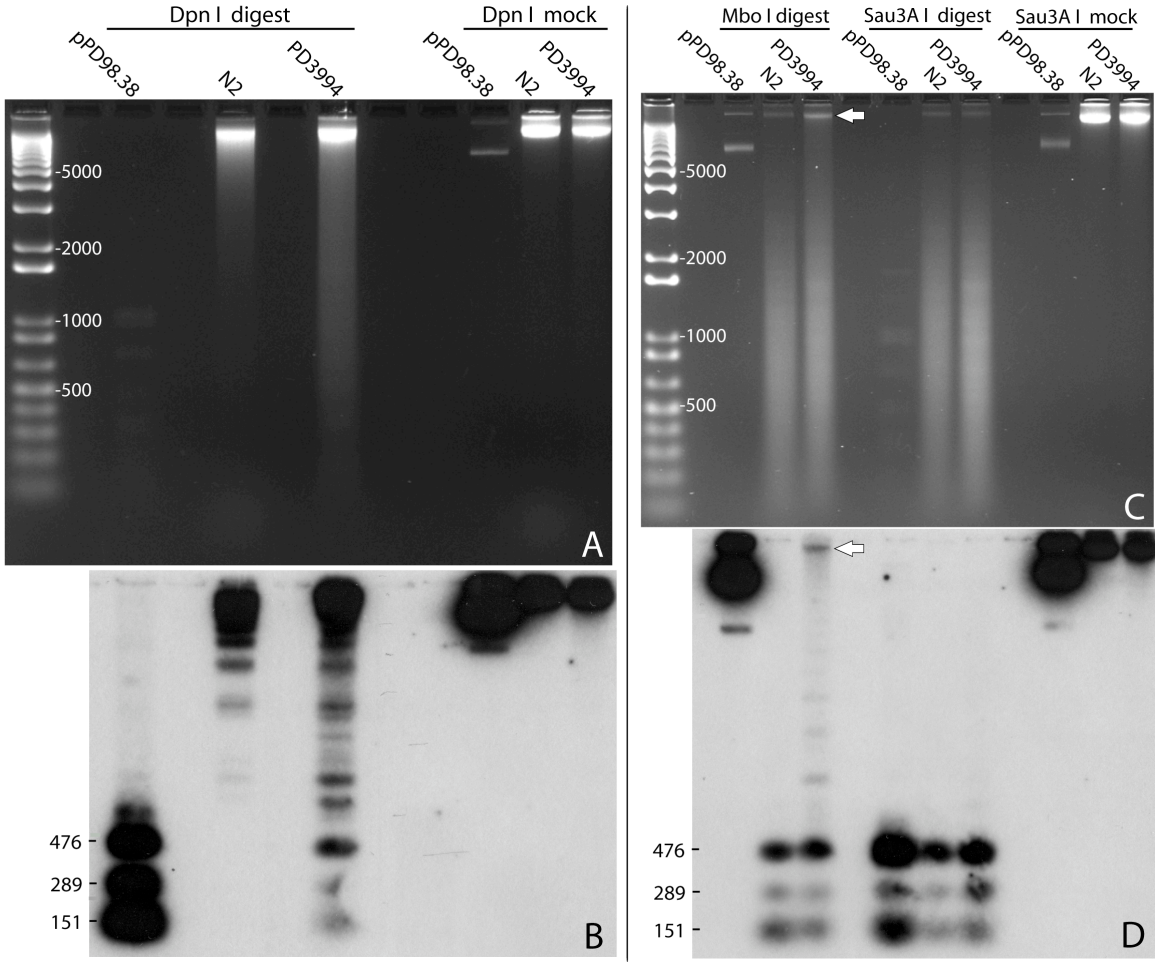


Figure 4.5. Southern blot analysis of PD5122 genomic DNA digested with *Dpn I*, *Mbo I*, or *Sau3A I*. (A, C, E) Ethidium bromide stained agarose gels. (B, D, F) The corresponding Southern blots to the agarose gels. Plasmid pPD98.38 was propagated in *E. coli* wildtype for dam methyltransferase and was, therefore, fully methylated. Hence, pPD98.38 was a substrated for both *Dpn I* and *Sau3A I*, but not *Mbo I*. The probe was a 808bp oligonucleotide containing the *C. elegans* 5S rDNA/SL1 sequence.

Figure 4.5

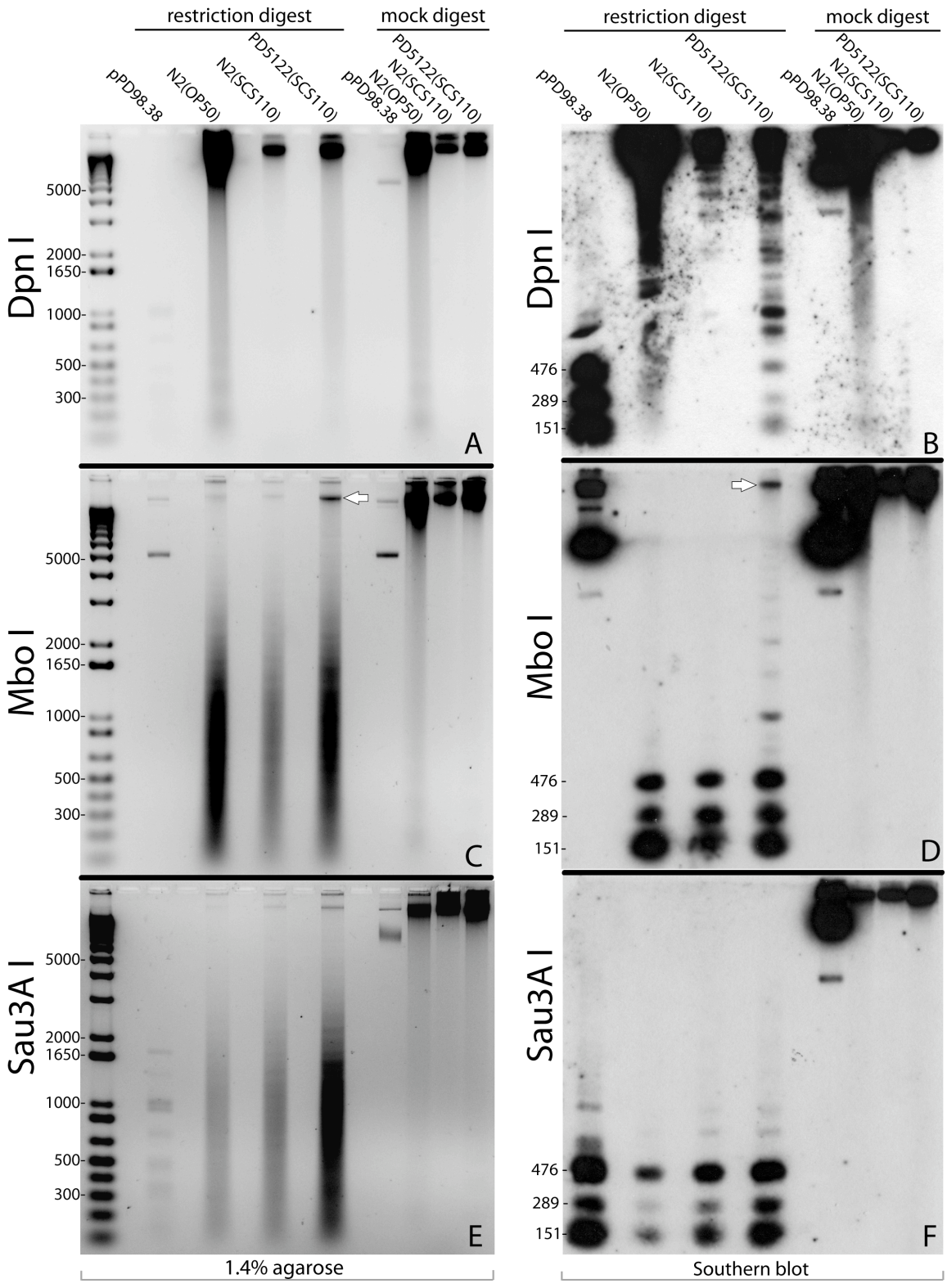
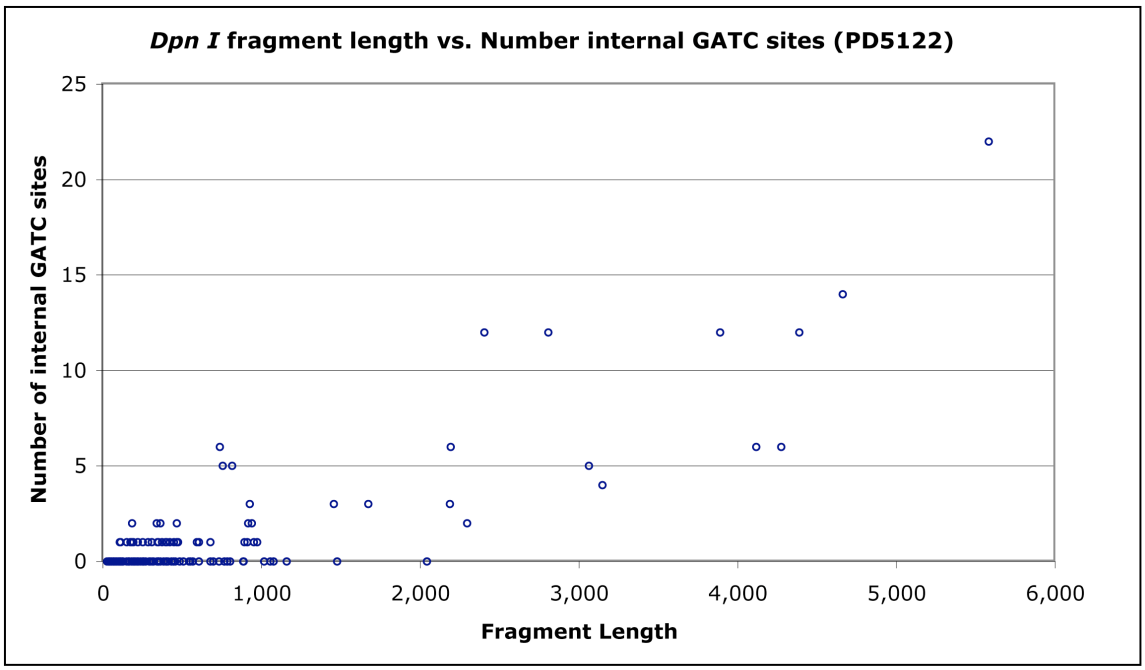

restriction digest
mock digest
pPD98.38
N2(OP50)
N2(SCS110)
PD5122(SCS110)
Dpn I
Mbo I
Sau3A I
5000
2000
1650
1000
500
300
476
289
151
A
B
C
D
E
F
1.4% agarose
Southern blot

Figure 4.6. *Dpn I* fragment length versus number of internal GATC sites. Only unique sequences in each transgenic line are plotted. Note that fragments with zero internal GATC sites are also plotted.



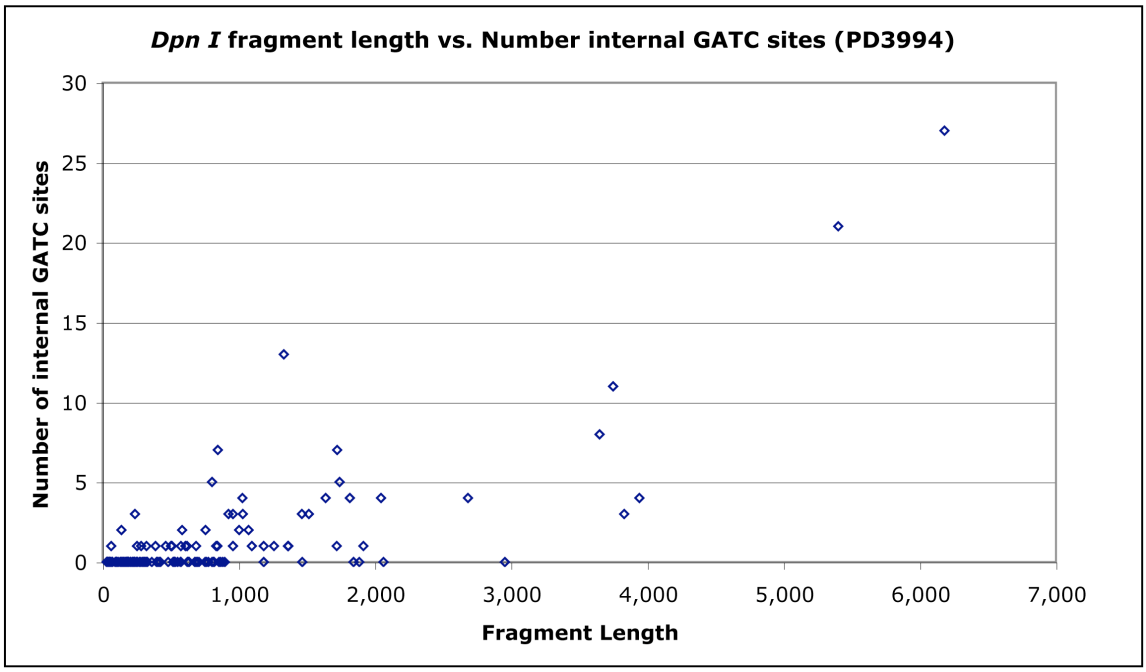


Figure 4.7. Size distribution of *Dpn I* fragments. The Y-axis is shown in log scale. The horizontal bar in each data set shows the average length for that data set.

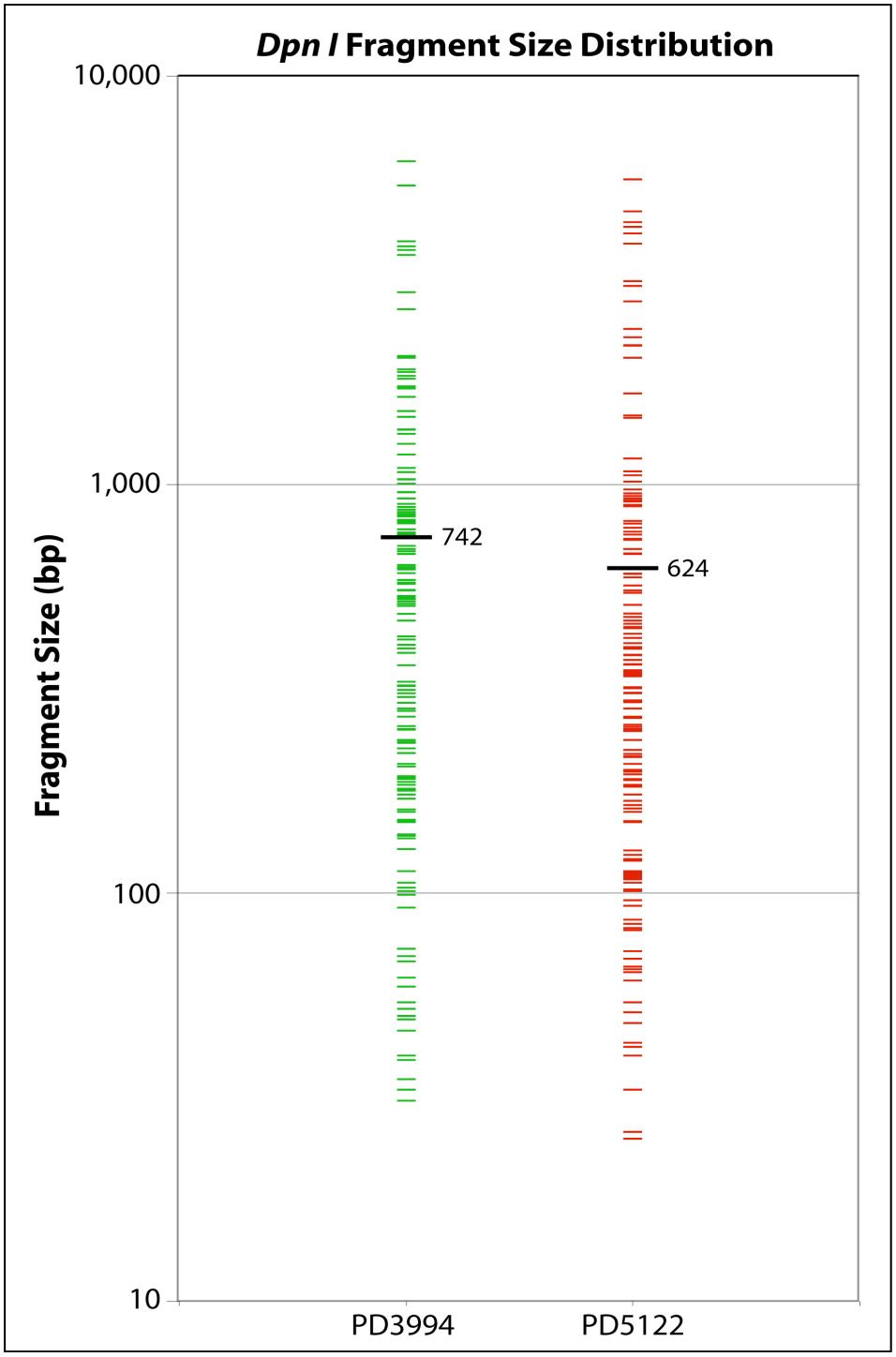
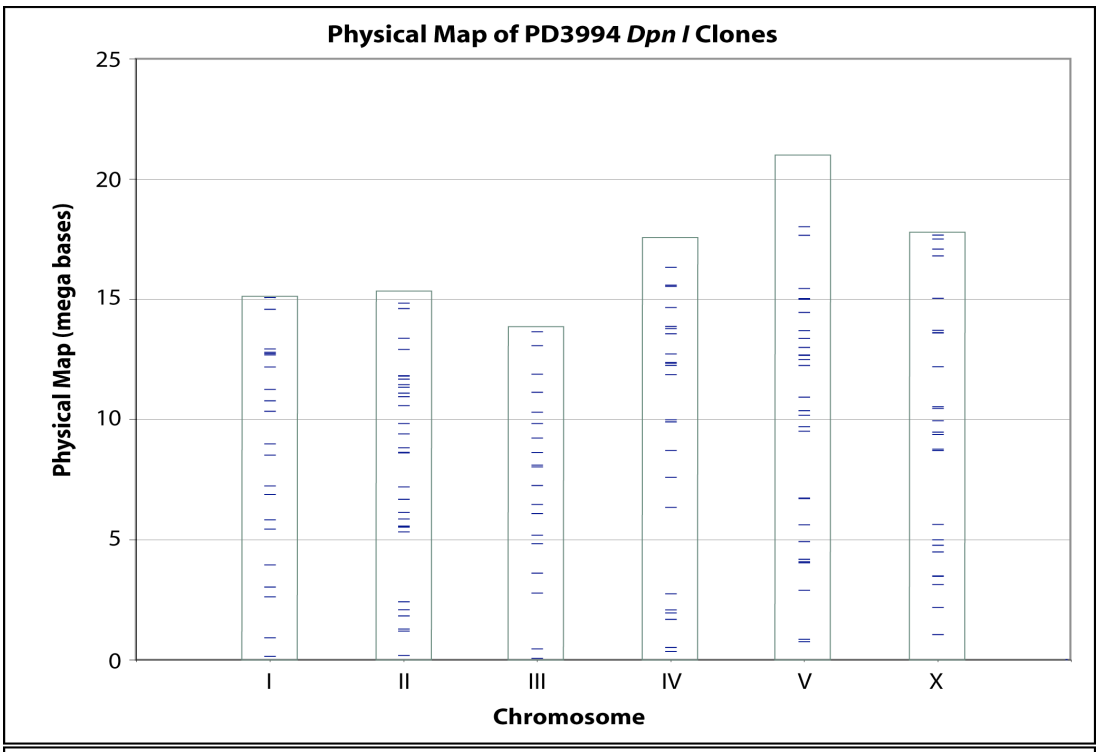


Figure 4.8. Physical map of *Dpn I* fragments. The vertical bars in each graph represent the physical size of each of the chromosomes. Notice that the Y-axis is in megabases.



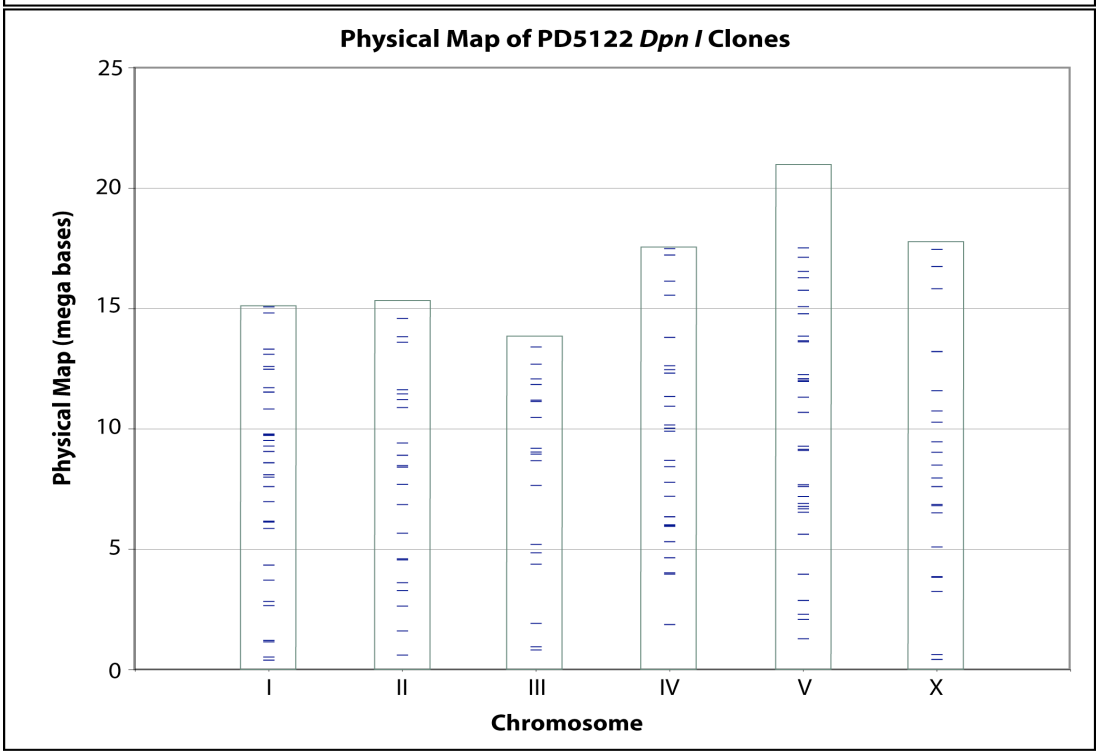
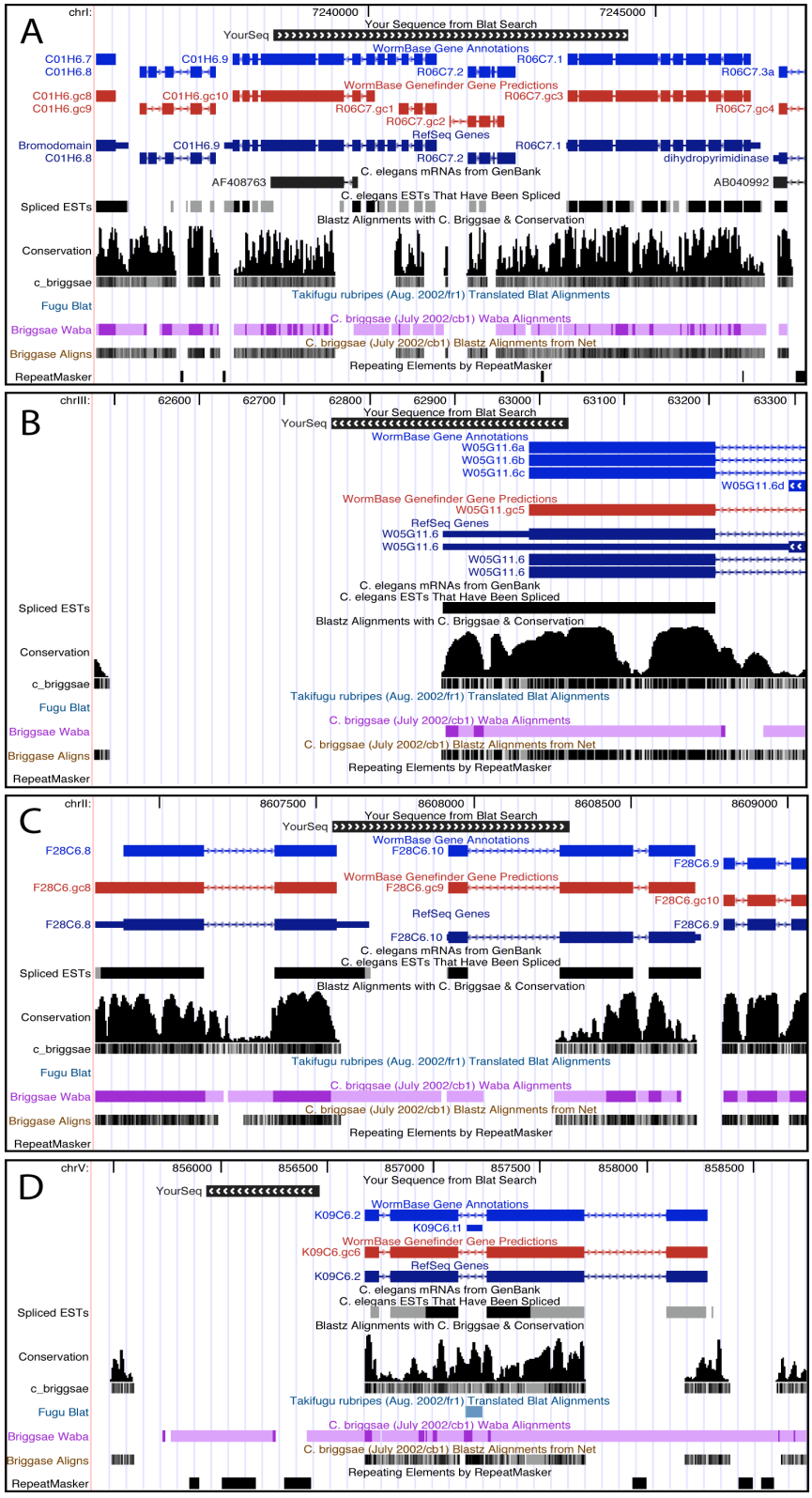


Figure 4.9. Rules for assigning scores to *Dpn I* fragment hits. (A) If a *Dpn I* fragment completely spans the gene (i.e. R06C7.2), that gene is given one point. Conversely, if the *Dpn I* fragment lies completely within a gene, that gene is given one point. (B) If a gene spans the *Dpn I* fragment on only one side (i.e. W05G11.6), that gene is given half a point. (C) In cases where each side of the *Dpn I* fragment spans a different gene, the gene with the greater overlap (i.e. F28C6.10) is given half a point. (D) If a *Dpn I* fragment does not span any gene, but a nearby gene lies within 1kb of the *Dpn I* fragment, that gene (i.e. K09C6.2) is given one point. If no genes are present within the 1kb limit on either side of the *Dpn I* fragment, then the *Dpn I* fragment is designated as 'not annotated' (i.e. intergenic). Screen captures taken from (http://genome.ucsc.edu/cgi-bin/hgBlat).

Figure 4.9


A
chrI:
7240000
7245000
Your Sequence from Blat Search
YourSeq
WormBase Gene Annotations
C01H6.7
C01H6.8
C01H6.9
R06C7.1
R06C7.2
R06C7.3a
WormBase Genefinder Gene Predictions
C01H6.gc8
C01H6.gc9
C01H6.gc10
R06C7.gc1
R06C7.gc2
R06C7.gc3
R06C7.gc4
RefSeq Genes
Bromodomain
C01H6.8
C01H6.9
R06C7.1
R06C7.2
dihydropyrimidinase
C. elegans mRNAs from GenBank
AF408763
AB040992
Spliced ESTs
C. elegans ESTs That Have Been Spliced
Blastz Alignments with C. Briggsae & Conservation
Conservation
c_briggsae
Takifugu rubripes (Aug. 2002/fr1) Translated Blat Alignments
Fugu Blat
C. briggsae (July 2002/cb1) Waba Alignments
Briggsae Waba
C. briggsae (July 2002/cb1) Blastz Alignments from Net
Briggase Aligns
Repeating Elements by RepeatMasker
RepeatMasker
B
chrIII:
62600
62700
62800
62900
63000
63100
63200
63300
Your Sequence from Blat Search
YourSeq
WormBase Gene Annotations
W05G11.6a
W05G11.6b
W05G11.6c
W05G11.6d
WormBase Genefinder Gene Predictions
W05G11.gc5
RefSeq Genes
W05G11.6
W05G11.6
W05G11.6
W05G11.6
C. elegans mRNAs from GenBank
C. elegans ESTs That Have Been Spliced
Spliced ESTs
Blastz Alignments with C. Briggsae & Conservation
Conservation
c_briggsae
Takifugu rubripes (Aug. 2002/fr1) Translated Blat Alignments
Fugu Blat
C. briggsae (July 2002/cb1) Waba Alignments
Briggsae Waba
C. briggsae (July 2002/cb1) Blastz Alignments from Net
Briggase Aligns
Repeating Elements by RepeatMasker
RepeatMasker
C
chrII:
8607500
8608000
8608500
8609000
Your Sequence from Blat Search
YourSeq
WormBase Gene Annotations
F28C6.8
F28C6.10
F28C6.9
WormBase Genefinder Gene Predictions
F28C6.gc8
F28C6.gc9
F28C6.gc10
RefSeq Genes
F28C6.8
F28C6.10
F28C6.9
C. elegans mRNAs from GenBank
C. elegans ESTs That Have Been Spliced
Spliced ESTs
Blastz Alignments with C. Briggsae & Conservation
Conservation
c_briggsae
Takifugu rubripes (Aug. 2002/fr1) Translated Blat Alignments
Fugu Blat
C. briggsae (July 2002/cb1) Waba Alignments
Briggsae Waba
C. briggsae (July 2002/cb1) Blastz Alignments from Net
Briggase Aligns
Repeating Elements by RepeatMasker
RepeatMasker
D
chrV:
856000
856500
857000
857500
858000
858500
Your Sequence from Blat Search
YourSeq
WormBase Gene Annotations
K09C6.2
K09C6.t1
WormBase Genefinder Gene Predictions
K09C6.gc6
RefSeq Genes
K09C6.2
C. elegans mRNAs from GenBank
C. elegans ESTs That Have Been Spliced
Spliced ESTs
Blastz Alignments with C. Briggsae & Conservation
Conservation
c_briggsae
Takifugu rubripes (Aug. 2002/fr1) Translated Blat Alignments
Fugu Blat
C. briggsae (July 2002/cb1) Waba Alignments
Briggsae Waba
C. briggsae (July 2002/cb1) Blastz Alignments from Net
Briggase Aligns
Repeating Elements by RepeatMasker
RepeatMasker

TABLES

Table 4.1. Summary of cloning results from PD3994 and PD5122. Each table lists the total, unique, and confirmed *Dpn I* clones each experiment. Of the confirmed *Dpn I* clones, the size range, average length, and average number of internal GATC sites are given for each chromosome.

PD3994 Cloning Summary

	I	II	III	IV	V	X	M	Total
total clones:	26	34	29	26	31	33	3	182
unique clones:	23	32	24	25	30	31	3	168
confirmed Dpn I clones:	22	31	20	24	27	26	3	153
unique clones — *Dpn I* clones — size range:	39-6,181	31-1,330	52-1,885	33-3,650	49-5,401	35-3,940	189-1,465	na
unique clones — *Dpn I* clones — avg. length of fragment:	971	432	470	742	1000	880	614	730
unique clones — *Dpn I* clones — avg. internal GATC per frag:	1.9	0.7	0.5	0.9	2.2	1.1	0.0	1.2

PD5122 Cloning Summary

	I	II	III	IV	V	X	M	Total
total clones:	34	22	22	32	35	25	4	174
unique clones:	33	21	20	32	34	23	4	167
confirmed Dpn I clones:	33	21	18	29	34	23	3	161
unique clones — *Dpn I* clones — size range:	43-4,662	42-2,187	25-5,581	40-2,294	51-4,117	26-4,388	48-308	na
unique clones — *Dpn I* clones — avg. length of fragment:	635	462	699	466	701	863	205	576
unique clones — *Dpn I* clones — avg. internal GATC per frag:	1.2	0.6	2.4	0.6	1.1	1.2	0.3	1.1

Table 4.2. *Dpn I* fragment hits from the PD3994 cloning experiment. Shown are unique, confirmed *Dpn I* fragments from the PD3994 cloning experiment. *Dpn I* fragments that are in intergenic regions, for which no annotation are given by the BLAT algorithm, are not shown. The physical map column gives the starting nucleotide of the *Dpn I* fragment.

Table 4.2

	PD3994 *Dpn I* Fragment Hits							
	gene	gene score	span	strand	physical map	frag length	internal GATC	repetitive?
Chromosome I	C32F10.6(flanking)	1	na	-	5,826,419	397	0	no
	F27C1.11	1	exon/intron	-	5,439,301	676	0	no
	F31C3.10	1	exon	-	15,068,878	191	0	yes
	F31C3.9	1	exon	+	15,073,791	39	0	no
	F32H2.5	1	exon/intron	-	8,984,085	178	0	no
	F46A8.3	1	exon/intron	+	11,250,522	1,464	3	yes
	F47B3.8	1	exon	+	3,947,018	329	0	no
	F56C11.1	1	intron	-	148,515	308	0	no
	K02B12.3	1	exon/intron	-	8,518,226	607	1	no
	R05D7.gc4(flanking)	1	exon/intron	+	12,182,659	3,830	3	no
	R06C7.2	1	exon/intron	+	7,238,349	6,181	27	no
	R13H8.1	1	intron	+	10,776,534	812	0	no
	T21G5.2	1	exon	+	6,879,427	50	0	no
	T26E3.2	1	exon/intron	+	12,687,319	624	0	no
	W02A11.3	1	exon/intron	+	12,746,982	1,097	1	no
	Y105E8A.24	1	exon	+	14,584,230	1,739	5	no
	Y18D10A.1	1	intron	-	12,797,076	752	0	yes
	Y18D10A.21	1	exon	-	12,937,684	136	0	no
	Y23H5A.3	1	exon/intron	-	2,620,506	844	1	no
	Y54E10BR.5	0.5	exon	+	3,031,057	170	0	no
	Y95B8A.12	1	intron	-	917,465	54	0	no
	ZC434.3	0.5	exon	+	10,343,592	881	0	no
	ZC434.5	0.5	exon	+	10,343,592	881	0	no
Chromosome II	B0491.5	1	exon/intron	-	11,344,558	504	1	no
	C09H10.3	1	exon/intron	+	11,099,337	552	0	no
	C14A4.1	1	exon/intron	+	10,580,284	151	0	no
	C17C3.2	0.5	exon/intron	-	5,568,071	322	1	no
	C29H12.1	1	exon/intron	+	6,132,141	1,072	2	no
	C47D12.2	1	exon	-	11,678,682	73	0	no
	E04F6.5	1	exon/intron	+	7,192,861	113	0	no
	F08B1.1	1	intron	+	5,325,620	776	0	no
	F19H8.4	1	exon	-	14,617,422	252	1	no
	F28C6.10	0.5	exon/intron	+	8,607,551	755	2	no
	F28C6.8	0.5	exon	+	8,607,551	755	2	no
	F28C6.9	1	exon/intron	+	8,610,238	149	0	no
	F42G2.4	1	intron	+	2,413,484	237	0	no
	F43G6.6	1	exon/intron	-	11,806,795	533	0	no
	F46C5.3(flanking)	1	na	-	8,825,296	187	0	no
	F53A10.2(flanking)	1	na	+	1,196,229	256	0	no
	F58G1.2	1	exon	+	12,917,389	302	0	no
	H43E16.1	1	exon/intron	+	6,678,363	835	1	no
	R03D7.7	1	exon	-	10,952,941	35	0	no
	R05H10.6	1	exon/intron	+	14,843,772	1,184	1	no
	T08E11.4	1	exon/intron	-	1,829,108	868	0	yes
	T16A1.7	1	exon	+	2,085,547	204	0	yes
	T22C8.7	1	exon/intron	-	8,632,651	686	0	no
	W02B12.1	1	exon/intron	+	11,449,394	628	0	no
	Y48C3A.12	1	intron	-	13,379,884	550	0	no
	Y51B9A.7	1	exon	-	9,400,307	31	0	no
	Y57G7A.12	1	intron	+	1,279,406	314	0	yes
	Y6D1A.1	1	exon/intron	+	11,821,009	1,330	13	yes
	ZK177.10(flanking)	1	na	+	5,524,406	184	0	no
	ZK938.1	1	exon/intron	+	9,830,288	204	0	no
Chromosome III	C05D10.4	1	exon/intron	+	6,089,418	92	0	no
	C13G5.1	1	exon/intron	+	8,622,968	425	0	no
	C14B9.4	0.5	exon	-	8,103,623	846	7	no
	C48D5.2	1	exon/intron	-	3,608,360	763	0	no
	F26A1.10	0.5	exon	+	4,830,227	128	0	no
	F44B9.7	1	exon/intron	+	8,025,228	524	0	no
	K10D2.1	1	exon/intron	+	5,189,126	897	0	no
	K11H3.4	1	exon/intron	+	9,833,059	417	0	no
	T07A5.1	1	exon/intron	+	10,304,789	1,885	0	no
	T12D8.1	1	exon	-	13,649,071	233	0	no
	T23G5.1	1	exon	-	9,228,026	52	0	no
	T26A5.9	1	exon/intron	+	6,463,396	283	1	no
	W05G11.6	0.5	exon	-	62,756	279	0	no
	Y39E4B.6	1	exon/intron	+	13,070,560	387	1	no
	Y48A6C.5	0.5	exon	-	11,135,542	575	0	yes
	Y55B1BR.5	1	exon/intron	-	455,112	150	0	no
	Y56A3A.5(flanking)	1	na	+	11,886,224	99	0	no
	Y71H2AM.19	1	exon	-	2,778,878	160	0	no

Table 4.2 (*continued*)

	PD3994 *Dpn I* Fragment Hits (*continued*)							
	gene	gene score	span	strand	physical map	frag length	internal GATC	repetitive?
Chromosome IV	B0513.5	1	exon/intron	-	13,872,928	804	0	no
	C29E6.2	1	exon	+	11,866,026	207	0	no
	C35D6.10	0.5	exon	+	16,332,680	321	0	no
	C47E12.6	1	exon/intron	+	9,987,900	322	0	no
	F19B6.4(flanking)	1	na	+	12,341,479	1,639	4	no
	F28D1.1	1	exon/intron	-	12,372,687	881	0	no
	F32B6.8(flanking)	1	na	+	9,898,347	361	0	no
	F52C12.2	1	exon	-	1,955,422	33	0	no
	H23L24.5	1	exon/intron	-	8,714,785	226	0	no
	R05C11.3	0.5	exon/intron	+	2,080,669	3,650	8	no
	T23F6.3	0.5	exon/intron	+	12,724,341	2,044	4	no
	T23F6.4	0.5	exon/intron	+	12,724,341	2,044	4	no
	W03G1.6	1	intron	-	512,775	819	0	no
	Y105C5A.4	1	exon/intron	+	15,589,800	2,063	0	yes
	Y2C2A.1	1	exon	-	7,595,176	139	2	yes
	Y45F10B.10	1	intron	+	13,565,861	46	0	no
	Y45F10D.3	1	exon/intron	+	13,779,573	292	0	no
	Y54G2A.12	0	exon	+	2,782,764	921	1	no
	Y57G11C.33	1	exon/intron	+	14,657,230	574	1	no
	Y66H1B.1	1	intron	+	353,442	527	0	yes
	Y73B6BL.15	1	exon	-	6,345,446	193	0	yes
	Y73F8A.35	1	exon/intron	+	15,537,191	956	1	no
	Y94H6A.1	1	exon/intron	+	2,745,268	220	0	no
Chromosome V	C02E7.5	0.5	exon/intron	+	4,920,424	958	3	yes
	C02E7.8	0.5	exon/intron	+	4,920,424	958	3	yes
	C05E4.3	1	exon	+	754,701	49	0	no
	C31B8.gc1	1	exon	+	2,895,397	70	0	no
	C49G7.7	1	exon/intron	-	4,037,163	405	0	no
	F11A3.1	0.5	exon/intron	+	9,512,616	1,722	7	no
	F11A3.3	0.5	exon/intron	+	9,512,616	1,722	7	no
	F14H8.6	1	exon/intron	-	15,033,898	233	0	no
	F15B9.1	1	exon/intron	+	12,999,329	1,030	3	no
	F23B12.6	1	exon/intron	-	14,453,905	5,401	21	no
	F58B4.6	0.5	exon	-	10,927,841	180	0	no
	F59A1.11	1	exon/intron	+	17,664,183	1,362	1	no
	K07C5.8	1	exon/intron	-	10,365,083	138	0	no
	K09C6.2(flanking)	1	na	-	855,933	530	0	no
	M03E7.2	1	exon/intron	-	5,620,508	270	0	no
	M04G12.1	1	intron	+	13,376,858	3,749	11	no
	T01C3.5	1	exon/intron	+	14,999,764	868	0	no
	T01D3.1	1	exon/intron	+	13,695,815	2,955	0	no
	T04H1.2	1	exon/intron	-	12,246,499	619	1	no
	T23F1.1	0.5	exon/intron	+	15,453,880	1,718	1	no
	T26H5.5	0.5	exon	+	15,453,880	1,718	1	no
	W02F12.3	0.5	exon/intron	+	6,706,827	584	2	no
	W02F12.4	0.5	exon/intron	+	6,706,827	584	2	no
	W04D2.3	0.5	exon	+	12,494,990	1,028	4	no
	W04D2.5	0.5	exon	+	12,494,990	1,028	4	no
	Y45G5AM.1	1	exon	+	4,187,809	237	3	no
	Y59A8B.4	1	exon	-	18,017,937	178	0	yes
	ZC487.1	1	exon/intron	+	6,730,996	158	0	no
	ZK287.8	1	intron	-	9,701,877	464	1	no
Chromosome X	C08A9.5	1	exon/intron	+	17,088,412	3,940	4	no
	C17G1.6	1	exon/intron	-	9,947,229	92	0	no
	C18B12.5	1	exon	-	15,043,089	103	0	no
	C39B10.1	1	exon/intron	-	10,460,519	925	3	no
	C39E6.6	1	exon/intron	+	4,767,265	510	1	no
	E01H11.1	1	intron	+	9,376,512	59	0	no
	F41G4.2	1	exon/intron	+	16,808,264	571	0	no
	F48B9.1	1	exon	-	2,180,977	106	0	no
	F53A9.3	0.5	exon/intron	+	8,707,025	1,514	3	no
	F53A9.4	0.5	exon	+	8,707,025	1,514	3	no
	F54B11.6	0.5	exon/intron	+	13,597,226	1,915	1	no
	K02D3.1(flanking)	1	na	+	13,705,538	1,183	0	no
	K03E6.3(flanking)	1	na	+	1,051,518	2,684	4	no
	K10B3.1	1	exon/intron	-	3,135,727	856	0	no
	M02D8.1(flanking)	1	na	+	8,763,355	174	0	no
	T23E7.2	1	exon/intron	-	17,674,064	635	0	no
	ZC8.6	1	exon/intron	+	4,992,155	1,814	4	no
	ZK899.8	1	intron	+	9,472,807	190	0	no
mito	MTCE.25	1	exon	+	7,439	189	0	no
	MTCE.34	1	exon	+	10,778	1,465	0	no

Table 4.3. *Dpn I* fragment hits from the PD5122 cloning experiment. Shown are unique, confirmed *Dpn I* fragments from the PD5122 cloning experiment. *Dpn I* fragments that are in intergenic regions, for which no annotations are given by BLAT algorithm, are not shown. The physical map column gives the starting nucleotide of the *Dpn I* fragment.

Table 4.3

	PD5122 *Dpn I* Fragment Hits							
	gene	gene score	span	strand	physical map	frag length	internal GATC	repetitive?
Chromosome I	C16C2.3	1	intron	+	9,729,732	108	0	no
	C30F12.3	0.5	exon/intron	+	6,978,309	604	1	no
	C35E7.6	1	exon	-	10,825,487	224	0	no
	C41G7.1	0.5	exon	+	9,519,063	296	0	no
	D2030.8	1	exon/intron	+	7,601,919	351	0	no
	F14B4.2(flanking)	1	na	-	9,286,913	168	0	no
	F18C12.2	1	exon	+	8,095,918	349	1	no
	F26E4.10	1	exon	+	9,790,019	108	1	no
	F30A10.10	0.5	exon	+	9,519,063	296	0	no
	F31C3.7(flanking)	1	na	-	15,072,590	121	0	no
	F32B4.4	1	exon	-	11,533,060	124	0	no
	F33D11.8(flanking)	1	na	+	5,863,872	1,053	0	no
	F33E2.7(flanking)	1	na	+	12,594,484	161	0	no
	F40E.5(flanking)	1	na	+	2,657,784	215	0	no
	H05L14.2	1	exon/intron	+	7,996,507	4,662	14	no
	K02A11.1	1	exon/intron	+	9,748,866	972	1	no
	R05D11.3	1	exon/intron	+	8,596,243	4,275	6	no
	R119.5	1	intron	+	392,290	43	0	no
	T04D3.1(flanking)	1	na	+	13,312,849	400	0	no
	T09B4.6(flanking)	1	na	-	6,165,085	195	0	no
	T12F5.4	1	exon	-	3,716,015	184	2	no
	T22H2.6	1	intron	+	11,711,364	110	0	no
	T27A3.1	0.5	exon/intron	-	6,130,715	483	0	no
	T27F6.1(flanking)	1	na	-	12,477,338	106	1	yes
	W06D4.1	1	exon/intron	-	9,059,206	255	0	no
	Y26D4A.14	1	exon	+	13,102,883	216	0	no
	Y48G8AL.1	0.5	exon/intron	+	1,211,430	1,476	0	no
	Y48G8AL.10	1	intron	+	1,149,117	84	0	yes
	Y54E5B.gc1	1	intron	+	14,813,807	421	1	no
	Y65B4BL.6(flanking)	1	na	+	524,745	2,806	12	no
	ZK973.6	1	exon	-	4,343,548	150	1	yes
Chromosome II	B0457.1	1	exon	-	8,906,458	69	0	no
	B0495.5	1	exon	+	7,696,767	252	0	no
	C01G12.8(flanking)	1	na	+	14,592,588	550	0	no
	C04A2.7c(flanking)	1	na	-	6,858,598	318	0	no
	C08E3.3	1	exon	+	1,604,697	66	0	no
	C25H3.15	1	exon/intron	+	5,665,691	237	0	no
	F02E11.1	1	exon/intron	+	3,280,207	346	1	no
	F07A11.6	1	exon	+	11,625,701	48	0	no
	F08G2.10	0.5	exon	+	13,829,359	1,157	0	no
	F08G2.4	0.5	exon	+	13,829,359	1,157	0	no
	F22E5.9(flanking)	1	na	-	2,637,836	102	0	no
	F35D11.5	1	intron	+	4,602,051	383	0	no
	F44F4.1	1	exon/intron	-	10,882,684	343	1	no
	F49E12.6(flanking)	1	na	+	8,412,429	82	0	no
	M03A1.3	1	exon/intron	+	4,559,075	2,187	3	no
	M05D6.1	0.5	exon	-	8,475,138	317	0	no
	M176.1(flanking)	1	na	+	9,412,153	182	0	no
	T06D8.1	1	exon/intron	+	11,215,479	283	1	yes
	W02B12.3	1	intron	+	11,454,573	42	0	no
	Y48E1B.13	1	exon/intron	-	13,602,484	372	1	no
	Y49F6A.5	1	exon/intron	+	3,608,973	736	6	no
Chromosome III	B0284.1	1	exon	+	4,378,415	120	0	no
	B0412.2	1	exon	-	813,450	106	0	no
	B0523.5	1	exon	-	8,678,866	783	0	no
	B0524.1(flanking)	1	na	-	1,917,427	182	1	yes
	D2045.9	1	exon/intron	+	10,476,357	951	1	no
	F11F1.7	1	intron	-	13,404,918	113	0	no
	F26A1.14	0.5	exon	-	4,849,698	124	0	no
	K10D2.6	1	exon/intron	-	5,198,038	464	1	no
	R107.1	1	exon	-	9,039,603	108	1	no
	T12B5.11	1	exon/intron	+	948,039	2,404	12	yes
	Y111B2A.19	1	intron	-	12,690,230	464	2	no
	Y47D3A.6	1	intron	-	11,193,621	112	1	no
	Y48A6C.5	1	exon	-	11,135,441	101	0	no
	Y56A3A.1	1	exon	-	11,845,140	25	0	no
	Y75B8A.1	1	exon/intron	-	12,074,515	362	2	yes
	ZC84.6	1	exon/intron	+	9,197,546	158	0	no
	ZK643.3	0.5	exon/intron	+	8,949,523	5,581	22	no
	ZK643.5	0.5	exon/intron	+	8,949,523	5,581	22	no
	ZK783.4	0.5	na	-	7,649,899	431	0	no

Table 4.3 (*continued*)

	PD5122 *Dpn I* Fragment Hits (*continued*)							
	gene	gene score	span	strand	physical map	frag length	internal GATC	repetitive?
Chromosome IV	4R79.1	1	exon/intron	-	17,488,972	127	0	no
	B0350.2	1	exon	+	5,988,952	695	0	no
	C01C7.1	1	exon/intron	-	12,624,033	543	0	no
	C25G4.8	1	exon/intron	+	12,462,193	592	1	no
	C33H5.9	1	exon	+	7,779,848	350	0	no
	C46G7.2	1	exon/intron	-	6,016,626	909	1	no
	F08B4.1	1	exon/intron	-	8,694,239	914	2	no
	F29B9.8	1	exon/intron	+	4,648,221	215	0	no
	F32B6.8	0.5	exon	+	9,902,457	93	0	no
	F44D12.5	0.5	exon	+	10,029,078	754	5	no
	F55A8.2	1	intron	+	1,868,385	400	1	no
	K09E10.3(flanking)	1	na	+	12,319,016	190	1	yes
	R13A1.4	1	exon/intron	-	7,200,834	885	0	no
	T01G1.1	1	exon/intron	-	11,345,807	346	0	no
	T14G10.1	1	exon/intron	-	10,162,641	339	2	no
	T26A8.4	1	exon	-	8,429,493	81	0	no
	W02C12.3	1	exon/intron	+	4,017,054	113	0	no
	Y105C5A.1(flanking)	1	na	+	15,554,689	72	0	no
	Y116A8B.4	1	exon/intron	+	17,223,019	676	1	no
	Y46C8AR.2	0.5	exon	+	3,964,488	566	0	no
	Y73B6BL.14	0.5	exon/intron	-	6,342,282	448	1	no
	Y73B6BL.gc18	1	intron	-	6,354,512	164	0	no
	ZK354.12	1	exon	+	5,316,621	2,294	2	yes
Chromosome V	B0213.11	1	intron	+	3,960,258	112	0	no
	B0213.14	1	intron	-	3,965,262	350	0	no
	C01G10.9	1	exon/intron	+	15,083,215	3,889	12	no
	C12D5.3(flanking)	1	na	-	7,680,479	474	1	no
	C50E3.11	1	intron	-	7,606,251	51	0	no
	F09F3.4	1	exon/intron	-	13,853,597	814	5	no
	F09G2.3	1	intron	+	7,190,190	732	0	yes
	F21C10.12(flanking)	1	na	+	9,101,353	937	2	no
	F29F11.6	1	exon/intron	-	10,685,197	349	0	no
	F37B4.10	1	exon	+	2,876,490	219	1	no
	F47G9.3	0.5	exon/intron	+	11,317,413	4,117	6	no
	F47G9.4	0.5	exon/intron	+	11,317,413	4,117	6	no
	F53F4.8	0.5	exon	-	13,620,160	268	0	no
	F57A10.1	1	exon	+	15,761,874	396	1	yes
	K03D7.11(flanking)	1	na	+	17,522,084	198	0	no
	K04A8.6	1	exon	+	6,536,464	149	0	no
	K11C4.4	1	exon	-	6,901,359	54	0	no
	K11G9.1	1	exon/intron	-	6,680,994	507	0	yes
	M03E7.2	1	exon/intron	-	5,620,508	270	0	no
	R07B7.10	1	exon/intron	+	12,085,882	3,063	5	no
	T04H1.4(flanking)	1	na	-	12,255,428	802	0	no
	T13F3.1	1	exon	-	16,279,988	249	0	no
	Y19D10B.2	1	exon/intron	-	2,301,783	86	0	no
	Y32B12B.2	1	exon/intron	+	16,543,015	363	0	no
	Y50E8A.9	1	exon	-	14,781,068	109	0	no
	ZC404.7	1	intron	-	6,776,482	207	0	no
	ZK1005.1	1	exon	+	1,282,871	82	0	no
	ZK218.12	1	na	+	17,132,534	293	0	yes
	ZK682.4	0.5	exon/intron	+	9,283,235	399	0	no
Chromosome X	C10E2.4	0.5	exon/intron	+	16,750,759	1,455	3	no
	C10E2.5	0.5	exon/intron	+	16,750,759	1,455	3	no
	C14F5.4	1	exon	+	7,958,159	48	0	no
	C36E6.3	1	exon	+	17,459,444	65	0	no
	C42D8.5	1	exon/intron	+	5,094,087	382	0	no
	C44E12.3	1	exon/intron	+	7,603,313	309	0	no
	C53B7.7	0.5	exon/intron	+	6,860,062	3,147	4	no
	F13C5.6(flanking)	1	na	+	622,107	2,192	6	no
	F22A3.2	1	exon	+	6,512,518	26	0	no
	F40F4.8	1	na	+	3,246,609	341	0	no
	F41E7.1	1	exon/intron	+	10,277,835	4,388	12	no
	F52D10.1	1	exon	+	11,585,860	33	0	no
	M03A8.2	0.5	exon	+	6,808,380	766	0	no
	R07A4.1	1	exon/intron	-	10,741,610	319	0	no
	R12H7.1	1	intron	-	13,205,595	605	0	no
	R12H7.4	1	exon/intron	-	13,218,187	678	0	no
	T01B10.5	1	exon/intron	-	8,497,637	174	1	no
	T25B6.7	1	exon/intron	+	9,030,711	341	0	no
	Y7A5A.8	1	intron	-	15,823,834	456	0	no
	ZK1193.1	1	exon	-	421,748	64	0	no
	ZK899.6(flanking)	0.5	na	+	9,465,019	2,042	0	no
mito	MTCE.23	1	exon	-	5,756	258	0	no
	MTCE.25	0.5	na	-	7,439	308	1	no
	MTCE.26	1	exon	-	7,896	48	0	no

Table 4.4. SAGE hits from the PD3994 cloning experiment. For each gene, we assigned a score ("weight") based on its proximity to the cloned *Dpn I* fragment (see Figure 4.9 for assignment rules). Shown in Tables 4.2 and 4.3 under the "gene score" heading are the weights for each gene. We then determined the weighted "hit frequency" for that gene in each tissue. This was done by multiplying the gene score by the number of hits for that gene. Finally, we determined the total weighted average for a particular tissue, based on all the weighted hits for that tissue. The formula for the weighted average is given by $x_w = \sum(w_i * x_i) / \sum(w_i)$, where w_i is the weight and x_i is the number of hits for that gene for a particular tissue. The weighted standard error is given by $(\sigma^2)_w / \sum(w_i)$, where $(\sigma^2)_w$ is the weighted variance and is given by $(\sigma^2)_w = \sum\left[w_i(x_i - x_w)^2\right] / \left[\sum(w_i) - 1\right]$

Table 4.4

	PD3994 SAGE Hits														
	gene	SW032	SWN22	SWEG1	SWEM1	SW031	SW030	SW028	SW023	SW033	SW034	SW035	SW037	SW038	SW039	SW040
Chromosome I	C32F10.6(flanking)	7	15	0	0	0	1	0	1	1	1	3	3	0	1	0
	F27C1.11	0	0	0	0	0	0	2	1	0	0	1	3	1	1	0
	F32H2.5	221	63	4	10	15	113	30	16	20	48	86	58	99	4	12
	F47B3.8	1	0	0	0	0	1	2	2	1	3	3	3	4	1	0
	F56C11.1	0	1	0	0	1	0	0	0	1	1	2	0	1	0	0
	K02B12.3	17	9	2	3	7	7	6	1	15	8	10	7	3	3	4
	R06C7.2	1	0	0	0	0	0	0	0	0	2	0	2	0	0	3
	R13H8.1	5	3	1	2	1	0	4	6	2	8	11	5	6	0	0
	T21G5.2	1	5	0	1	0	1	0	0	0	0	0	0	1	0	0
	T26E3.2	0	9	5	4	30	12	5	1	18	17	5	10	11	5	0
	W02A11.3	15	8	3	3	11	14	45	10	4	7	13	7	9	6	3
	Y105E8A.24	3	0	1	0	2	0	6	6	0	5	18	11	13	1	0
	Y106G6D.7(flanking)	1	2	1	0	3	2	2	4	4	2	12	6	6	2	3
	Y18D10A.1	5	0	0	1	0	0	1	4	2	0	0	1	10	2	1
	Y23H5A.3	2	9	0	0	3	1	1	1	1	2	2	1	1	5	13
	Y95B8A.12	0	0	0	0	0	0	5	0	0	0	0	2	0	0	0
	ZC434.3	0	2	0	0	0	0	0	1	0	1	0	1	0	0	0
	ZC434.5	2	1	8	1	3	5	4	1	14	2	2	3	3	2	12
	avg. weighted score:	17.4	7.7	1.3	1.5	4.5	9.5	6.8	3.1	4.5	6.5	9.7	7.2	10.0	1.9	2.6
Chromosome II	B0491.5	47	15	16	12	21	31	11	2	41	16	7	12	7	0	12
	C09H10.3	12	7	4	3	13	18	12	2	29	15	15	11	11	67	19
	C14A4.1	37	14	19	8	31	50	15	4	51	18	16	6	22	22	17
	C15F1.2	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0
	C29H12.1	3	4	0	0	1	0	1	0	0	1	0	3	0	1	4
	C47D12.2	14	14	0	0	5	10	0	2	4	7	8	5	2	0	2
	E04F6.5	3	15	8	5	3	13	10	5	12	11	10	10	4	1	2
	F08B1.1	18	17	8	1	4	15	21	1	11	2	3	5	59	0	2
	F19H8.4	3	0	0	0	0	0	0	0	0	0	3	4	0	1	0
	F28C6.10	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	F28C6.8	5	6	5	0	2	1	0	0	6	2	1	0	2	0	1
	F46C5.3(flanking)	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0
	F53A10.2(flanking)	3	5	0	0	5	5	3	0	1	9	4	4	3	6	2
	F58G1.2	3	2	3	0	2	4	2	0	7	0	0	0	6	1	1
	H43E16.1	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0
	R03D7.7	5	1	0	0	0	0	0	0	0	1	0	1	0	0	0
	T08E11.4	1	5	1	0	1	0	0	2	1	3	0	1	3	1	0
	T16A1.7	0	0	3	0	0	0	0	0	1	0	1	0	0	0	0
	T22C8.7	0	1	0	0	0	1	0	1	0	1	0	1	0	0	0
	W02B12.1	1	2	0	0	1	0	0	0	0	0	0	0	1	0	0
	Y48C3A.12	3	1	0	0	1	2	0	2	0	1	1	3	1	7	2
	Y6D1A.1	20	14	1	1	3	1	0	0	1	8	1	4	1	0	3
	ZK938.1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
	avg. weighted score:	8.381	5.7	3.2	1.4	4.4	7.2	3.6	1.0	7.7	4.5	3.4	3.4	5.9	5.1	3.2
Chromosome III	C05D10.4	4	12.0	2.0	1.0	3.0	5.0	7.0	0.0	8.0	10.0	9.0	8.0	15.0	9.0	0.0
	C13G5.1	29	9	3	0	3	8	0	1	3	6	4	1	2	2	0
	C14B9.4	42	11	8	5	7	7	22	7	5	18	7	20	11	45	27
	C34E10.8	1	4	1	1	1	1	0	2	1	1	0	1	2	0	0
	C48D5.2	2	7	4	6	8	7	19	6	14	19	11	21	12	16	0
	F44B9.7	0	0	0	0	1	0	0	1	1	1	0	0	0	1	0
	K02F3.6	2	2	1	0	2	4	3	0	2	2	5	4	3	0	0
	K10D2.1	22	8	0	0	1	2	2	1	3	3	7	3	4	5	6
	K11H3.4	1	0	0	0	0	0	0	0	0	1	0	1	3	1	2
	R74.1	17	21	6	3	4	13	14	4	16	5	11	7	13	2	10
	T12D8.1	43	21	4	5	13	14	25	4	7	13	13	17	23	25	3
	T23G5.1	56	26	1	1	4	4	1	5	5	6	3	14	5	1	39
	T26A5.9	45	76	8	49	111	47	2	0	88	4	12	5	2	13	87
	W05G11.6	18	69	1	20	55	28	2	0	11	18	4	21	38	0	0
	Y39E4B.6	16	9	3	4	10	8	3	1	11	13	5	13	4	16	0
	Y48A6C.5	4	0	0	1	0	2	0	2	2	0	8	1	0	0	0
	Y56A3A.5(flanking)	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
	Y71H2AM.19	0	0	0	0	0	2	1	0	1	1	1	2	1	11	0
	avg. weighted score:	18.5	15.4	2.2	5.9	13.7	8.6	5.3	1.7	11.1	7.0	5.5	7.9	7.1	9.1	11.1

purified oocytes (SW032)
N2 embryos (SWN22)
gut cells (SWEG1)
muscle cells (SWEM1)
muscle cells (SW031)
hypodermal cells (SW030)
pan-neural cells (SW028)
ciliated neuron (SW023)
pharynx cells (SW033)
AFD neuron (SW034)
pharyngeal marginal (SW035)
ACER neuron (SW037)
punc-4::GFP cells (SW038)
pharyngeal gland cells (SW039)
dissected gonad (SW040)

Table 4.4 (*continued*)

		PD3994 SAGE Hits (*continued*)														
	gene	SW032	SWN22	SWEG1	SWEM1	SW031	SW030	SW028	SW023	SW033	SW034	SW035	SW037	SW038	SW039	SW040
Chromosome IV	B0513.5	1	1	0	0	1	1	2	2	0	2	2	1	2	0	0
	C29E6.2	2	0	0	0	1	2	1	3	4	3	4	3	2	1	0
	C47E12.6	0	1	0	0	0	8	1	1	1	1	0	2	0	0	0
	F28D1.1	11	8	3	2	5	9	3	3	4	8	9	12	11	40	6
	F32B6.8(flanking)	8	11	2	1	3	11	9	9	8	8	17	9	7	12	10
	F52C12.2	7	9	3	2	4	6	4	3	18	3	2	5	5	3	10
	H23L24.5	2	0	2	0	0	4	0	1	1	0	0	1	34	0	0
	R05C11.3	1	4	0	1	1	2	1	4	1	2	9	1	12	18	0
	T23F6.4	4	7	1	1	2	5	3	1	3	8	12	13	9	1	0
	W03G1.6	13	7	1	0	1	4	3	6	3	3	2	2	3	0	8
	Y2C2A.1	0	2	0	0	0	0	1	2	0	1	2	8	24	0	0
	Y45F10B.10	1	1	0	0	0	0	3	1	1	1	3	1	3	4	0
	Y45F10D.3	21	11	6	1	6	4	6	1	6	9	13	6	4	53	4
	Y54G2A.12	1	0	0	1	0	0	0	1	0	0	0	0	0	0	2
	Y57G11C.33	7	14	1	0	3	4	4	0	15	6	6	4	5	21	8
	Y73F8A.35	0	1	0	0	0	0	0	1	2	0	2	0	2	21	0
	Y94H6A.1	1	0	0	0	0	0	0	0	0	0	1	2	1	0	0
	avg. weighted score:	5.1	4.8	1.2	0.5	1.7	3.8	2.6	2.4	4.3	3.3	4.9	4.2	7.6	11.0	3.1
Chromosome V	C03A7.1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
	C05E4.3	0	0	0	1	4	1	2	2	0	0	2	3	0	9	0
	C49G7.7	7	0	12	0	0	17	7	3	8	1	1	0	155	2	0
	F11A3.1	6	8	2	2	4	11	1	0	5	2	1	5	0	0	0
	F14H8.6	1	1	0	0	0	2	0	1	0	1	0	0	29	0	0
	F23B12.6	10	7	1	1	3	8	4	5	5	7	8	6	4	1	8
	F58B4.6	0	2	0	0	1	6	12	2	4	0	4	12	13	0	0
	F59A1.11	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0
	K07C5.8	18	11	1	2	3	6	0	0	7	5	5	5	4	3	7
	M03E7.2	0	12	2	1	2	1	1	1	6	1	3	11	11	0	0
	M04G12.1	0	0	3	0	1	4	1	0	0	2	1	1	2	1	0
	T01D3.1	0	0	2	1	1	2	4	7	1	5	6	3	5	1	0
	T04H1.2	9	3	1	0	3	11	2	2	9	5	3	1	14	7	11
	W02F12.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
	W02F12.4	3	8	1	2	1	0	3	3	4	1	0	4	7	1	1
	W04D2.3	1	1	1	0	3	0	1	1	6	1	3	1	3	0	2
	W04D2.5	13	16	26	12	21	26	7	7	44	24	9	9	8	32	11
	Y45G5AM.1	1	0	0	1	0	0	0	0	0	0	0	1	1	2	0
	ZC487.1	0	1	0	0	1	3	3	1	7	3	8	1	3	0	0
	ZK287.8	10	2	1	1	1	2	0	2	2	5	3	3	1	5	0
	avg. weighted score:	4.2	3.4	2.4	1.0	2.1	4.9	2.3	1.9	4.8	3.1	3.0	3.2	15.3	3.0	2.2
Chromosome X	C02H7.1	1	1	0	1	0	0	3	3	1	2	3	7	3	0	0
	C17G1.6	1	0	0	0	0	0	0	0	0	0	0	0	3	0	0
	C18B12.5	4	2	1	0	1	2	0	1	1	2	2	1	1	0	0
	C39B10.1	0	1	0	0	0	0	1	0	0	1	1	0	1	0	0
	C39E6.6	0	0	0	0	0	0	3	1	0	0	0	1	0	0	0
	C56E10.1	2	1	2	0	0	0	1	0	2	4	6	3	6	5	0
	E01H11.1	1	3	0	0	1	0	2	1	1	2	2	4	5	0	0
	F41G4.2	18	24	2	12	25	17	9	7	31	34	19	27	19	2	1
	F53A9.3	0	11	2	0	3	3	0	1	17	3	1	4	1	1	0
	F53A9.4	0	1	0	0	1	2	2	2	4	3	15	3	4	0	0
	F54B11.6	13	17	1	2	14	12	14	5	41	28	26	28	10	4	0
	K02D3.1(flanking)	3	3	0	0	2	1	0	0	0	0	1	1	1	1	0
	K03E6.3(flanking)	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0
	K10B3.1	13	26	0	1	7	26	0	0	11	0	0	1	0	0	0
	M02D8.1(flanking)	0	0	0	3	15	0	0	0	6	1	2	0	1	0	0
	R03E1.2	12	24	3	3	8	12	2	0	9	7	1	3	8	4	2
	T23E7.2	3	3	0	4	5	1	1	1	2	6	6	1	5	2	0
	ZC8.6	0	0	0	0	0	0	1	0	1	1	1	1	0	6	1
	ZK899.8	10	18	3	4	2	6	11	7	14	5	11	10	18	18	2
	avg. weighted score:	4.1	6.6	0.5	1.7	4.6	4.2	2.6	1.5	6.8	4.8	4.6	4.4	4.3	2.2	0.3
	total weighted avg (cloning):	**9.5**	**7.0**	**1.9**	**1.9**	**4.9**	**6.4**	**3.8**	**1.9**	**6.5**	**4.8**	**5.1**	**4.9**	**8.4**	**5.2**	**3.6**
	total avg. (SAGE db):	**7.5**	**6.7**	**2.7**	**1.8**	**4.8**	**6.0**	**3.8**	**2.6**	**7.3**	**4.3**	**5.4**	**4.1**	**5.7**	**4.3**	**4.9**
	weighted standard error:	6.3	1.5	0.1	0.3	1.8	2.1	0.5	0.1	1.7	0.6	1.0	0.6	4.2	1.3	1.1

purified oocytes (SW032)
N2 embryos (SWN22)
gut cells (SWEG1)
muscle cells (SWEM1)
muscle cells (SW031)
hypodermal cells (SW030)
pan-neural cells (SW028)
ciliated neuron (SW023)
pharynx cells (SW033)
AFD neuron (SW034)
pharyngeal marginal (SW035)
ACER neuron (SW037)
punc-4::GFP cells (SW038)
pharyngeal gland cells (SW039)
dissected gonad (SW040)

Table 4.5. SAGE hits from the PD5122 cloning experiment. For each gene, we assigned a score ("weight") based on its proximity to the cloned *Dpn I* fragment (see Figure 4.9 for assignment rules). Shown in Tables 4.2 and 4.3 under the "gene score" heading are the weights for each gene. We then determined the weighted "hit frequency" for that gene in each tissue. This was done by multiplying the gene score by the number of hits for that gene. Finally, we determined the total weighted average for a particular tissue, based on all the weighted hits for that tissue. The formula for the weighted average is given by $x_w = \sum(w_i * x_i) / \sum(w_i)$, where w_i is the weight and x_i is the number of hits for that gene for a particular tissue. The weighted standard error is given by $(\sigma^2)_w / \sum(w_i)$, where $(\sigma^2)_w$ is the weighted variance and is given by $(\sigma^2)_w = \sum\left[w_i(x_i - x_w)^2\right] / \left[\sum(w_i) - 1\right]$

Table 4.5

	PD5122 SAGE Hits														
	gene	SW032	SWN22	SWEG1	SWEM1	SW031	SW030	SW028	SW023	SW033	SW034	SW035	SW037	SW038	SW039	SW040
Chromosome I	C16C2.3	2	10	1	0	3	1	1	0	3	0	5	2	5	0	1
	C30F12.3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
	C35E7.6	0	0	0	0	0	0	0	0	0	2	0	0	2	0	0
	C41G7.1	11	15	1	1	6	10	4	9	6	8	24	8	2	0	10
	D2030.8	12	7	0	0	0	0	1	1	2	3	1	3	2	0	5
	F14B4.2(flanking)	7	14	4	2	2	20	6	3	15	6	10	3	21	0	29
	F18C12.2	8	4	2	1	0	0	1	2	9	2	5	2	5	2	1
	F26E4.10	10	3	0	0	1	1	3	0	2	1	3	3	5	1	0
	F30A10.10	34	12	5	2	2	6	5	22	10	10	17	18	10	2	0
	F32B4.4	4	4	2	0	1	3	0	0	5	5	3	0	1	7	1
	F32B4.4	4	4	2	0	1	3	0	0	5	5	3	0	1	7	1
	F33D11.8(flanking)	2	0	0	1	1	0	0	0	4	0	0	0	6	0	0
	H05L14.2	0	0	0	0	0	0	0	1	0	0	2	2	1	0	0
	K02A11.1	18	31	4	4	4	11	7	7	22	8	15	5	16	0	1
	R05D11.3	31	11	5	2	11	8	5	1	10	2	5	1	6	1	42
	R119.5	5	3	2	0	3	2	7	1	4	3	3	2	2	36	0
	T04D3.1(flanking)	3	2	0	0	0	0	0	0	0	1	1	1	0	0	1
	T09B4.6(flanking)	34	49	4	10	8	19	0	0	21	0	0	0	0	0	0
	T12F5.4	4	5	2	2	0	7	10	8	5	13	8	5	21	15	2
	T22H2.6	2	18	13	1	3	3	2	0	6	4	1	2	1	1	0
	T27A3.1	6	8	3	2	6	6	3	2	7	2	1	1	6	0	1
	W06D4.1	21	31	9	0	4	61	7	4	8	9	15	12	9	0	1
	Y26D4A.14	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
	Y48G8AL.1	9	4	1	0	0	2	1	2	0	5	3	3	5	27	5
	Y48G8AL.10	1	0	0	0	0	0	1	0	0	4	0	0	1	1	0
	Y65B4BL.6(flanking)	0	15	5	2	7	12	3	3	22	6	4	13	5	1	0
	ZK973.6	55	8	4	4	6	13	56	35	12	50	56	50	191	219	1
	avg. weighted score:	10.3	9.7	2.6	1.3	2.5	7.2	4.8	3.4	6.8	5.6	6.6	4.9	12.8	12.5	3.8
Chromosome II	B0457.1	13	29	5	3	5	8	12	1	12	12	7	10	13	2	7
	B0495.5	19	9	1	1	3	2	4	0	3	3	6	4	5	1	6
	C08E3.3	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
	F02E11.1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
	F07A11.6	57	15	3	2	3	12	169	20	7	41	52	43	35	51	0
	F08G2.4	8	11	6	3	11	5	1	0	16	3	4	3	11	1	1
	F35D11.5	3	7	6	2	10	3	4	1	18	3	3	2	6	1	13
	F44F4.1	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0
	F49E12.6(flanking)	28	18	9	3	10	3	7	1	14	10	10	9	1	0	0
	M03A1.3	2	2	0	0	0	2	2	3	1	0	3	1	2	2	0
	T06D8.1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0
	W02B12.3	9	18	1	1	6	7	13	8	22	16	21	3	8	49	19
	Y48E1B.13	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0
	avg. weighted score:	11.0	8.4	2.2	1.1	3.4	3.2	17.0	2.7	6.8	6.9	8.5	6.1	6.3	8.5	3.6
Chromosome III	B0284.1	0	0	0	0	0	2	0	1	0	0	0	0	0	0	0
	B0412.2	0	0	0	0	0	0	0	6	0	0	1	0	38	0	0
	B0523.5	16	10	4	4	4	3	3	1	7	7	14	6	6	3	3
	B0524.1(flanking)	1	0	0	0	0	0	2	3	0	2	3	1	2	1	0
	D2045.9	5	3	0	1	6	2	2	0	3	0	8	2	1	1	0
	F11F1.7	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
	K10D2.6	20	12	8	0	1	9	8	3	6	6	11	1	5	0	10
	R107.1	98	7	6	0	4	9	3	0	4	3	2	2	1	4	0
	T12B5.11	5	4	2	0	1	0	3	1	4	1	5	3	5	0	0
	Y111B2A.19	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1
	Y47D3A.6	1	2	1	3	2	1	1	0	4	0	0	1	1	1	0
	Y48A6C.5	4	0	0	1	0	2	0	2	2	0	8	1	0	0	0
	Y48A6C.5	4	0	0	1	0	2	0	2	2	0	8	1	0	0	0
	Y56A3A.1	15	6	2	1	3	9	9	6	4	6	10	11	9	21	14
	Y75B8A.1	0	2	0	0	2	3	0	0	0	0	0	0	1	3	0
	ZC84.6	0	3	0	1	1	1	2	3	3	2	7	1	1	9	0
	ZK643.3	0	0	0	0	0	0	0	1	0	0	0	0	4	0	0
	ZK643.5	19	19	8	5	14	22	23	4	33	15	16	13	31	3	0
	ZK783.4	27	15	4	3	3	7	10	4	8	14	14	14	19	3	0
	avg. weighted score:	11.0	3.8	1.7	1.0	1.9	3.3	2.8	1.9	3.4	2.4	5.3	2.5	5.5	2.6	1.6

purified oocytes (SW032)
N2 embryos (SWN22)
gut cells (SWEG1)
muscle cells (SWEM1)
muscle cells (SW031)
hypodermal cells (SW030)
pan-neural cells (SW028)
ciliated neuron (SW023)
pharynx cells (SW033)
AFD neuron (SW034)
pharyngeal marginal (SW035)
ACER neuron (SW037)
punc-4::GFP cells (SW038)
pharyngeal gland cells (SW039)
dissected gonad (SW040)

Table 4.5 (*continued*)

	PD5122 SAGE Hits (*continued*)															
	gene	SW032	SWN22	SWEG1	SWEM1	SW031	SW030	SW028	SW023	SW033	SW034	SW035	SW037	SW038	SW039	SW040
Chromosome IV	4R79.1	0	3	2	0	1	1	1	1	1	2	6	4	9	0	0
	B0350.2	43	25	16	36	82	34	93	39	46	78	79	90	139	105	1
	C01C7.1	2	9	1	11	12	3	12	6	2	6	12	4	13	0	1
	C33H5.9	5	5	1	3	4	6	0	1	4	1	8	1	2	1	2
	C46G7.2	4	12	1	11	55	6	0	0	8	3	0	2	0	0	1
	F08B4.1	1	7	1	1	1	4	3	5	2	1	1	2	6	0	0
	F29B9.8	1	3	0	6	4	2	0	2	0	1	1	1	0	0	1
	F29B9.8	1	3	0	6	4	2	0	2	0	1	1	1	0	0	1
	F32B6.8	8	11	2	1	3	11	9	9	8	8	17	9	7	12	10
	F55A8.2	21	19	0	3	6	18	16	10	10	9	15	9	14	5	19
	R13A1.4	0	1	0	0	0	1	0	0	0	0	1	2	3	2	0
	T01G1.1	14	4	4	3	5	7	10	13	15	12	25	14	36	3	1
	T14G10.1	19	32	2	1	6	17	12	7	26	3	7	13	10	4	11
	T26A8.4	16	8	3	4	4	17	8	8	16	3	5	9	23	0	4
	W02C12.3	7	4	2	5	5	5	2	5	3	2	2	5	9	12	1
	Y105C5A.1(flanking)	4	5	0	0	1	0	3	0	1	2	3	4	1	1	0
	Y73B6BL.14	0	0	0	0	0	0	0	0	0	0	1	0	0	0	2
	avg. weighted score:	8.9	9.1	2.1	5.7	12.0	8.0	10.3	6.5	8.6	8.0	10.9	10.3	16.8	8.7	3.1
Chromosome V	B0213.11	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
	B0213.14	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	C01G10.9	2	0	1	0	0	2	1	1	2	5	2	0	1	9	0
	C12D5.3(flanking)	0	0	0	0	0	0	0	0	0	0	0	1	4	1	0
	C50E3.11	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
	F09G2.3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	F21C10.12(flanking)	0	0	0	0	0	0	0	0	2	0	1	0	0	0	0
	F29F11.6	9	6	3	1	3	13	1	2	7	4	8	0	5	2	13
	F37B4.10	4	7	3	1	2	6	11	13	13	0	60	7	6	2	2
	F47G9.3	0	0	0	0	0	0	1	0	2	1	1	1	1	1	0
	F47G9.4	0	0	0	1	1	0	0	0	0	1	4	2	1	0	0
	F53F4.8	2	0	2	1	0	0	1	1	0	0	0	0	6	0	0
	K04A8.6	6	39	11	6	11	29	9	4	26	20	6	17	19	2	0
	K11C4.4	235	10	8	2	4	10	3	1	2	7	5	5	1	3	0
	K11G9.1	0	0	0	0	1	1	0	0	0	0	0	1	5	1	0
	M03E7.2	0	12	2	1	2	1	1	1	6	1	3	11	11	0	0
	R07B7.10	2	1	1	0	0	2	1	2	2	1	5	0	0	10	1
	T04H1.4(flanking)	55	13	0	2	3	3	8	1	5	6	12	6	13	6	0
	T13F3.1	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Y19D10B.2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	Y32B12B.2	5	6	1	1	1	4	2	0	1	2	2	1	3	17	5
	ZK1005.1	8	5	8	0	0	1	1	1	0	1	1	2	12	1	0
	ZK682.4	0	0	1	0	0	0	1	5	4	1	8	6	0	0	0
	avg. weighted score:	16.1	4.8	1.9	0.7	1.3	3.4	1.9	1.4	3.3	2.3	5.3	2.6	4.0	2.6	1.0
Chromosome X	C10E2.4	1	0	0	1	0	0	0	0	2	0	0	0	1	0	0
	C14F5.4	1	5	1	1	3	2	1	1	13	5	3	4	2	0	0
	C36E6.3	24	74	5	39	165	18	10	19	35	21	13	27	98	2	2
	C42D8.5	4	11	1	2	1	32	3	2	6	4	5	11	4	0	0
	C44E12.3	0	2	0	0	0	1	12	1	4	2	2	1	2	0	1
	F13C5.6(flanking)	5	18	2	11	18	2	1	9	1	15	17	8	21	0	0
	F22A3.2	2	23	6	41	99	15	2	1	10	3	1	2	0	0	0
	F40F4.8	0	0	1	0	0	0	0	0	0	0	0	0	2	0	0
	F41E7.1	0	0	0	0	0	0	0	1	0	2	5	1	2	0	0
	F52D10.1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	M03A8.2	6	9	2	3	1	7	11	14	10	24	38	19	13	39	0
	R07A4.1	0	0	0	0	1	0	1	0	2	0	1	0	1	1	0
	R12H7.1	1	7	3	3	10	2	12	4	7	4	4	4	16	1	0
	R12H7.4	0	1	0	0	1	0	0	0	0	1	0	0	1	0	0
	T01B10.5	0	0	2	0	1	1	1	1	2	1	2	0	2	0	0
	T25B6.7	1	1	0	0	0	0	1	0	0	1	0	0	0	0	0
	ZK1193.1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	11
	ZK899.6(flanking)	12	8	0	1	3	1	0	0	1	1	0	0	0	0	0
	avg. weighted score:	3.1	9.1	1.3	6.0	18.2	4.7	3.0	2.8	5.2	4.3	4.4	4.1	9.6	1.4	0.9
total weighted avg (cloning):		**10.3**	**7.5**	**2.0**	**2.5**	**6.1**	**5.1**	**5.9**	**3.1**	**5.6**	**4.7**	**6.7**	**4.9**	**9.3**	**6.3**	**2.3**
total avg. (SAGE db):		**7.5**	**6.7**	**2.7**	**1.8**	**4.8**	**6.0**	**3.8**	**2.6**	**7.3**	**4.3**	**5.4**	**4.1**	**5.7**	**4.3**	**4.9**
weighted standard error:		6.4	1.2	0.1	0.4	3.9	0.7	3.4	0.4	0.6	1.0	1.4	1.2	5.7	5.5	0.3

purified oocytes (SW032)
N2 embryos (SWN22)
gut cells (SWEG1)
muscle cells (SWEM1)
muscle cells (SW031)
hypodermal cells (SW030)
pan-neural cells (SW028)
ciliated neuron (SW023)
pharynx cells (SW033)
AFD neuron (SW034)
pharyngeal marginal (SW035)
ACER neuron (SW037)
punc-4::GFP cells (SW038)
pharyngeal gland cells (SW039)
dissected gonad (SW040)

ACKNOWLEDGMENTS

The SAGE data were produced at the Michael Smith Genome Sciences Centre with funding from Genome Canada.

REFERENCES

ACHWAL, C. W., C. A. IYER and H. S. CHANDRA, 1983 Immunochemical evidence for the presence of 5mC, 6mA and 7mG in human, Drosophila and mealybug DNA. FEBS Lett **158:** 353-358.

ARDIZZI, J. P., and H. F. EPSTEIN, 1987 Immunochemical localization of myosin heavy chain isoforms and paramyosin in developmentally and structurally diverse muscle cell types of the nematode *Caenorhabditis elegans*. J Cell Biol **105:** 2763-2770.

BARNES, T. M., Y. KOHARA, A. COULSON and S. HEKIMI, 1995 Meiotic recombination, noncoding DNA and genomic organization in *Caenorhabditis elegans*. Genetics **141:** 159-179.

BARRAS, F., and M. G. MARINUS, 1989 The great GATC: DNA methylation in *E. coli*. Trends Genet **5:** 139-143.

BAYLIN, S. B., 2005 DNA methylation and gene silencing in cancer. Nat Clin Pract Oncol **2 Suppl 1:** S4-11.

BONGIORNI, S., O. CINTIO and G. PRANTERA, 1999 The relationship between DNA methylation and chromosome imprinting in the coccid Planococcus citri. Genetics **151:** 1471-1478.

BONGIORNI, S., and G. PRANTERA, 2003 Imprinted facultative heterochromatization in mealybugs. Genetica **117:** 271-279.

BOYE, E., M. G. MARINUS and A. LOBNER-OLESEN, 1992 Quantitation of Dam methyltransferase in Escherichia coli. J Bacteriol **174:** 1682-1685.

BRENNER, S., 1974 The genetics of *Caenorhabditis elegans*. Genetics **77:** 71-94.

BROWN, S. W., 1959 Lecanoid chromosome behavior in three more families of the Coccoidea (Homoptera). Chromosoma **10:** 278-300.

BROWN, S. W., 1961 Fracture and fusion of coccid chromosomes. Nature **191:** 1419-1420.

BROWN, S. W., and W. A. NELSEN-REES, 1961 Radiation analysis of the lecanoid genetic system. Genetics **42:** 510-523.

BROWN, S. W., and H. S. CHANDRA, 1977 *Chromosome imprinting and the differential regulation of homologous chromosomes*. Academic Press, New York.

CAMBARERI, E. B., M. J. SINGER and E. U. SELKER, 1991 Recurrence of repeat-induced point mutation (RIP) in Neurospora crassa. Genetics **127:** 699-710.

CONSORTIUM, 1998 Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science **282:** 2012-2018.

DEL GAUDIO, R., R. DI GIAIMO and G. GERACI, 1997 Genome methylation of the marine annelid worm *Chaetopterus variopedatus*: methylation of a CpG in an expressed H1 histone gene. FEBS Lett **417:** 48-52.

DEOBAGKAR, D. N., K. MURALIDHARAN, S. G. DEVARE, K. KALGHATGI and H. S. CHANDRA, 1982 The mealybug chromosome system I: Unusual methylated bases and dinucleotides in DNA of a Planococcus species. J Biosci (India) **4:** 513.

DIBB, N. J., I. N. MARUYAMA, M. KRAUSE and J. KARN, 1989 Sequence analysis of the complete *Caenorhabditis elegans* myosin heavy chain gene family. J Mol Biol **205:** 603-613.

EROVA, T. E., L. PILLAI, A. A. FADL, J. SHA, S. WANG *et al.*, 2006 DNA adenine methyltransferase influences the virulence of Aeromonas hydrophila. Infect Immun **74:** 410-424.

FALKER, S., M. A. SCHMIDT and G. HEUSIPP, 2005 DNA methylation in Yersinia enterocolitica: role of the DNA adenine methyltransferase in mismatch repair and regulation of virulence factors. Microbiology **151:** 2291-2299.

FEDOREYEVA, L. I., and B. F. VANYUSHIN, 2002 N(6)-Adenine DNA-methyltransferase in wheat seedlings. FEBS Lett **514:** 305-308.

FEIL, R., and S. KHOSLA, 1999 Genomic imprinting in mammals: an interplay between chromatin and DNA methylation? Trends Genet **15:** 431-435.

FIELD, L. M., F. LYKO, M. MANDRIOLI and G. PRANTERA, 2004 DNA methylation in insects. Insect Mol Biol **13:** 109-115.

FINNEGAN, E. J., R. I. BRETTELL and E. S. DENNIS, 1993 The role of DNA methylation in the regulation of plant gene expression. Exs **64:** 218-261.

FIRE, A., W. KELLY, M. HSU, S. XU, J. AHNN *et al.*, 1998 The Uses of Green Fluorescent Protein in *Caenorhabditis elegans*., pp. 153-168 in *Green Fluorescent Protein: Properties, Applications, and Protocols*., edited by M. C. A. S. KANE. Wiley-Liss, Inc., New York.

FIRE, A., R. ALCAZAR and F. TAN, 2006 Unusual DNA Structures Associated With Germline Genetic Activity in *Caenorhabditis elegans*. Genetics **173:** 1259-1273.

GRACE GOLL, M., and T. H. BESTOR, 2005 Eukaryotic cytosine methyltransferases. Annu Rev Biochem **74:** 481-514.

GRANATO, M., H. SCHNABEL and R. SCHNABEL, 1994 *pha-1*, a selectable marker for gene transfer in *C. elegans*. Nucleic Acids Res **22:** 1762-1763.

GREENWALD, I., A. COULSON, J. SULSTON and J. PRIESS, 1987 Correlation of the physical and genetic maps in the lin-12 region of *Caenorhabditis elegans*. Nucleic Acids Res **15:** 2295-2307.

GUTIERREZ, A., and R. J. SOMMER, 2004 Evolution of dnmt-2 and mbd-2-like genes in the free-living nematodes Pristionchus pacificus, Caenorhabditis elegans and Caenorhabditis briggsae. Nucleic Acids Res **32:** 6388-6396.

HARFE, B. D., A. VAZ GOMES, C. KENYON, J. LIU, M. KRAUSE *et al.*, 1998 Analysis of a *Caenorhabditis elegans* Twist homolog identifies conserved and divergent aspects of mesodermal patterning. Genes Dev **12:** 2623-2635.

HEITHOFF, D. M., R. L. SINSHEIMER, D. A. LOW and M. J. MAHAN, 1999 An essential role for DNA adenine methylation in bacterial virulence. Science **284:** 967-970.

HINRICHS, A. S., D. KAROLCHIK, R. BAERTSCH, G. P. BARBER, G. BEJERANO *et al.*, 2006 The UCSC Genome Browser Database: update 2006. Nucleic Acids Res **34:** D590-598.

HONDA, S., and H. F. EPSTEIN, 1990 Modulation of muscle gene expression in *Caenorhabditis elegans*: differential levels of transcripts, mRNAs, and polypeptides for thick filament proteins during nematode development. Proc Natl Acad Sci U S A **87:** 876-880.

KAPOOR, A., F. AGIUS and J. K. ZHU, 2005 Preventing transcriptional gene silencing by active DNA demethylation. FEBS Lett **579:** 5889-5898.

KARRER, K. M., and S. STEIN-GAVENS, 1990 Constancy of adenine methylation in Tetrahymena macronuclear DNA. J Protozool **37:** 409-414.

KARRER, K. M., and T. A. VANNULAND, 1998 Position effect takes precedence over target sequence in determination of adenine methylation patterns in the nuclear genome of a eukaryote, Tetrahymena thermophila. Nucleic Acids Res **26:** 4566-4573.

KARRER, K. M., and T. A. VANNULAND, 2002 Methylation of adenine in the nuclear DNA of Tetrahymena is internucleosomal and independent of histone H1. Nucleic Acids Res **30:** 1364-1370.

KENT, W. J., 2002 BLAT--the BLAST-like alignment tool. Genome Res **12:** 656-664.

KHOSLA, S., G. MENDIRATTA and V. BRAHMACHARI, 2006 Genomic imprinting in the mealybugs. Cytogenet Genome Res **113:** 41-52.

KRAUSE, M., and D. HIRSH, 1987 A trans-spliced leader sequence on actin mRNA in *C. elegans*. Cell **49:** 753-761.

KUDRIASHOVA, I. B., M. D. KIRNOS and B. F. VANIUSHIN, 1976 [DNA-methylase activities from animal mitochondria and nuclei: different specificity of DNA methylation]. Biokhimiia **41:** 1968-1977.

LYKO, F., B. H. RAMSAHOYE, H. KASHEVSKY, M. TUDOR, M. A. MASTRANGELO *et al.*, 1999 Mammalian (cytosine-5) methyltransferases cause genomic DNA methylation and lethality in Drosophila. Nat Genet **23:** 363-366.

MARINUS, M. G., and N. R. MORRIS, 1973 Isolation of deoxyribonucleic acid methylase mutants of *Escherichia coli* K-12. J Bacteriol **114:** 1143-1150.

MCKAY, S. J., R. JOHNSEN, J. KHATTRA, J. ASANO, D. L. BAILLIE *et al.*, 2003 Gene expression profiling of cells, tissues, and developmental stages of the nematode *C. elegans*. Cold Spring Harb Symp Quant Biol **68:** 159-169.

MEYER, P., I. HEIDMANN and I. NIEDENHOF, 1993 Differences in DNA-methylation are associated with a paramutation phenomenon in transgenic petunia. Plant J **4:** 89-100.

MILLER, D. M., F. E. STOCKDALE and J. KARN, 1986 Immunological identification of the genes encoding the four myosin heavy chain isoforms of *Caenorhabditis elegans*. Proc Natl Acad Sci U S A **83:** 2305-2309.

NELSON, D. W., and B. M. HONDA, 1989 Two highly conserved transcribed regions in the 5S DNA repeats of the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae*. Nucleic Acids Res **17:** 8657-8667.

NEUHUBER, F., Y. D. PARK, A. J. MATZKE and M. A. MATZKE, 1994 Susceptibility of transgene loci to homology-dependent gene silencing. Mol Gen Genet **244:** 230-241.

OKKEMA, P. G., S. W. HARRISON, V. PLUNGER, A. ARYANA and A. FIRE, 1993 Sequence requirements for myosin gene expression and regulation in *Caenorhabditis elegans*. Genetics **135:** 385-404.

OZA, J. P., J. B. YEH and N. O. REICH, 2005 DNA methylation modulates Salmonella enterica serovar Typhimurium virulence in Caenorhabditis elegans. FEMS Microbiol Lett **245:** 53-59.

PALMER, B. R., and M. G. MARINUS, 1994 The dam and dcm strains of *Escherichia coli*--a review. Gene **143:** 1-12.

PRADHAN, S., and P. O. ESTEVE, 2003 Mammalian DNA (cytosine-5) methyltransferases and their expression. Clin Immunol **109:** 6-16.

PRASAD, S. S., and D. L. BAILLIE, 1989 Evolutionarily conserved coding sequences in the dpy-20-unc-22 region of *Caenorhabditis elegans*. Genomics **5:** 185-198.

RUSSELL, D. W., and J. SAMBROOK (Editors), 2001 *Southern Hybridization of Radiolabeled Probes to Nucleic Acids Immobilized on Membranes*. CSHL Press, Cold Spring Harbor.

SCHNABEL, H., and R. SCHNABEL, 1990 An organ-specific differentiation gene, *pha-1*, from *Caenorhabditis elegans*. Science **250:** 686-688.

SCHRADER, F., and S. HUGHES-SCHRADER, 1931 Haploidy in metazoa. Q Rev Biol **6:** 411-438.

SIMPSON, V. J., T. E. JOHNSON and R. F. HAMMEN, 1986 Caenorhabditis elegans DNA does not contain 5-methylcytosine at any time during development or aging. Nucleic Acids Res **14:** 6711-6719.

STARR, T., A. M. HOWELL, J. MCDOWALL, K. PETERS and A. M. ROSE, 1989 Isolation and mapping of DNA probes within the linkage group I gene cluster of *Caenorhabditis elegans*. Genome **32:** 365-372.

TWEEDIE, S., J. CHARLTON, V. CLARK and A. BIRD, 1997 Methylation of genomes and genes at the invertebrate-vertebrate boundary. Mol Cell Biol **17:** 1469-1475.

URIG, S., H. GOWHER, A. HERMANN, C. BECK, M. FATEMI *et al.*, 2002 The *Escherichia coli* dam DNA methyltransferase modifies DNA in a highly processive reaction. J Mol Biol **319:** 1085-1096.

VAN DEN WYNGAERT, I., J. SPRENGEL, S. U. KASS and W. H. LUYTEN, 1998 Cloning and analysis of a novel human putative DNA methyltransferase. FEBS Lett **426:** 283-289.

WADE, P. A., 2001 Methyl CpG-binding proteins and transcriptional repression. Bioessays **23:** 1131-1137.

WATERSTON, R., C. MARTIN, M. CRAXTON, C. HUYNH, A. COULSON *et al*., 1992 A survey of expressed genes in *Caenorhabditis elegans*. Nat Genet **1:** 114-123.

WINES, D. R., P. B. TALBERT, D. V. CLARK and S. HENIKOFF, 1996 Introduction of a DNA methyltransferase into Drosophila to probe chromatin structure in vivo. Chromosoma **104:** 332-340.

WION, D., and J. CASADESUS, 2006 N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. Nat Rev Microbiol **4:** 183-192.

YODER, J. A., and T. H. BESTOR, 1998 A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. Hum Mol Genet **7:** 279-284.

CHAPTER 5

Conclusion and Perspectives

The work described in this dissertation has been an attempt to understand epigenetic mechanisms in *C. elegans*. Ultimately, to understand how epigenetics operate, we must understand chromatin. It is through modulation of chromatin structures that organisms bring about changes in gene activity that can be stably inherited through mitosis or meiosis.

As recently as the mid 1980's, chromatin was but a definition: the total sum of DNA and the proteins that bind to it. Chromatin was thought of as static structures that had little role in mediating gene activity. Transcription factors ruled. Beginning in the early 1990's, pioneering work by numerous investigators led to an explosion in understanding of chromatin structure and its involvement in gene regulation. Today, the definition of chromatin is still the sum of DNA and all the proteins that bind to it. But the meaning has changed. It is a complex, dynamic regulatory network involving DNA, proteins, and RNA. And, it has emerged as an active field of investigation in biology.

Understanding the initiation of gene silencing: The continuing search for *cis*-acting "context" and *trans*-acting factors that trigger silencing of foreign DNA

We began our work with an attempt to identify factors that modulate transgene activities in *C. elegans*. In *C. elegans* as in many other species, introduction of transgenes into the organism often results in silencing of the reporter construct. Such a response is thought to be a defensive system against invasion by selfish genetic elements such as viruses and transposable elements, and there are some experimental evidence to support this view (ARAVIN *et al.* 2001; GRISHOK *et al.* 2000; HIROCHIKA *et al.* 2000; JENSEN *et al.* 1999; JEONG BR *et al.* 2002; KETTING *et al.* 1999; NOLAN *et al.* 2005; SIJEN and PLASTERK 2003; VASTENHOUW *et al.* 2003; WOODHOUSE *et al.* 2006; WU-SCHARF *et al.*

2000). How does an organism distinguish between its own DNA and foreign DNA? Cells employ various chromatin modifications such as DNA methylation, histone modification, and polycomb/trithorax group proteins to mark certain sequences of DNA. Chromatin modifications confer extra information to DNA beyond its primary sequence, such as parent-of-origin, self versus foreign DNA, and transcriptional competency. However, DNA introduced by researchers is often naked DNA that is devoid of such marks. There must be other contexts that cells employ to distinguish self from non-self.

One commonality that transgenes share is the tendency for unit transgenes to form long, tandemly repeated arrays consisting of many units. Paradoxically, it is generally the case that the higher the number of copies, the less activity is seen for the transgene. This phenomenon, called repeat-induced gene silencing (RIGS), has been observed in many organisms including *C. elegans* (HSIEH and FIRE 2000; MELLO *et al.* 1991; STINCHCOMB *et al.* 1985), plants (ASSAAD *et al.* 1993; DAY *et al.* 2000; YE and SIGNER 1996), mammals (GARRICK *et al.* 1998; MCBURNEY *et al.* 2002; PALMITER and BRINSTER 1986), fungi (SELKER 1990; SELKER 1997), and *Drosophila* (DORER and HENIKOFF 1994). In plants, mammals, and fungi, highly repetitive transgenes are often targeted for DNA methylation. Thus, DNA methylation may mark transgenes as foreign and recruit other factors to repress expression of the transgene. Whatever the mechanism(s) of recognition and silencing of transgenes, the fate of such transgenes is usually heterochromatization that renders them transcriptionally incompetent. This can be assayed in a number of ways, including low or absence of reporter gene activity, direct microscopic observation of condensed chromatin, low mRNA level, positive staining of antibodies directed against de-activating chromatin modifications, and/or resistance to nuclease digestion.

We note, however, that DNA methylation is not necessarily the only step leading to the silencing of repetitive transgenes. For example, *Drosophila* has very low DNA methylation activity (KUNERT *et al*. 2003), and *C. elegans* does not have any detectable DNA methylation activity; yet highly repetitive transgenes are just as efficiently silenced in these organisms (as assayed by the same methods above). It is possible that *C. elegans* and *Drosophila* (and other organisms as well), employ RNA-mediated transcriptional silencing. In plants, RNA has been shown to direct methylation of homologous DNA sequences (PELISSIER and WASSENEGGER 2000). This may be the initiating step leading to heterochromatization of target sequences. In *C. elegans* and *Drosophila,* aberrant RNAs (siRNAs, dsRNA, anti-sense RNAs) can be generated from aberrant transcription of a repetitive transgene (FIRE *et al.* 1991). These aberrant RNAs may initiate silencing of the transgene through a post-transcriptional mechanism (i.e. RNAi), as RNAi has been proposed (KETTING *et al*. 1999; TABARA *et al*. 1999) to be able to mediate transcriptional silencing in *C. elegans* and *Drosophila* (GRISHOK *et al*. 2005; PAL-BHADRA *et al*. 2004).

How do cells distinguish between foreign highly repeated DNA and endogenous DNA with repetitive character (i.e. heat shock, ribosomal, and tRNA genes), and only target the former for silencing? A very strong determinant of transgene activity in any organism is genomic position. Transgenes that integrate into heterochromatin are often heterochromatinized and silenced. However, transgenes flanked by insulator elements are protected from position effects and generally express well even when integrated into heterochromatin (GDULA *et al*. 1996; LABRADOR and CORCES 2002; LI *et al*. 2002; TAJIMA *et al*. 2006). Similar mechanisms may protect highly repetitive endogenous gene clusters from silencing. In *Drosophila*, the heat shock gene cluster is kept

transcriptionally active by the two insulator elements *scs* and *scs*' (UDVARDY *et al*. 1985). The chicken β-globin locus is similarly protected from silencing by a flanking barrier element (CHUNG *et al*. 1993; CHUNG *et al*. 1997). Interestingly, transcriptional competency of the ribosomal gene cluster does not appear to be similarly regulated by boundary elements. Instead, it appears that a fraction of genes in the cluster are kept in an open chromatin (transcriptionally competent) form (GRUMMT and PIKAARD 2003; SANTORO *et al*. 2002).

Alternatively, one might postulate that *trans*-acting factors somehow act to silence high-copy transgenes but not low-copy arrays. It is difficult to envisage how molecular recognition of repeat character might be carried out. Perhaps the repeat nature might recruit *trans*-acting factors cooperatively, concentrating silencing activity to a localized region. Such a mechanism, however, would require at least some recognition of sequence specificity. Yet, most silencing factors are able to act on unrelated transgene arrays that do not share any sequence homology, suggesting a sequence-independent mechanism. An alternative view might be that *trans*-acting factors play a secondary role, rather than a primary role, in the silencing of transgene arrays. In this scenario, repeat character somehow leads to initiation of a heterochromatic state in the transgene array. Guided by the histone code, *trans*-acting factors would land on appropriate sites on heterochromatin to maintain the silenced state. Heterochromatin protein 1 (HP1) and the Sir proteins in yeast appear to operate by such a mechanism.

One interesting thing to note here. *Neurospora* is the ideal system in which to study repeat-induced silencing. Numerous genetic screens have been carried out to analyze this process, and yet the (very interesting) results of these screens somehow don't

include anything with even a remote chance of being involved in the actual recognition event. This argues both for the challenges (and payoffs) associated with further genetic study of the recognition machinery.

Epigenetic memory and amnesia: maintenance and resetting of epigenetic states in diverse developmental contexts

In our work to understand transgene behavior in *C. elegans*, we have unexpectedly observed that we could confer a novel property to a set of transgenes. In particular, integration of the extra-chromosomal transgene array *ccEx3815*[*unc-54::gfp* + *pha-1*(+)] led to the acquisition of a parent-of-origin effect in two of the three integrated derivatives (*ccIn3852* and *ccIn3862*). Interestingly, a similar "simple" assay with another integrated derivative of *ccEx3815*, *ccIn3861*, did not yield detectable imprinting effects; while an unrelated *unc-119::gfp* fusion protein and *ccEx3815* itself exhibit a parent-of-origin effect. The magnitude of the imprinted effect in all these transgenes is modest, approximately two-fold from sperm transmission compared to oocyte transmission. We postulate that some (as yet unknown) chromatin configuration (size, repetitive character, complexity, integration site, etc.) may allow these transgenes to undergo differential chromatin modifications during spermatogenesis or oogenesis. Interestingly, we note that *ccEx3815* has not undergone any obvious structural rearrangements during the three independent integration events (Chapter 2 Part I, Figure 2.10), although we are not able to ascertain whether it has undergone modest changes in copy number.

Why would we observe imprinting for transgenes while similar processes appear rarely (if at all) for endogenous genes? A few investigators have advanced the view that

loci where selfish genetic elements have integrated serve either as "attractors" or "mediators" of epigenetic phenomena (LIPPMAN *et al.* 2004; LYON 2003; WHITELAW and MARTIN 2001). This hypothesis is supported by much experimental and observational evidence. It appears that sites of epigenetic processes are invariably associated with loci with high repetitive character and tandem repeats of a complex nature (head-to-tail or head-to-head arrangement of tandemly repeated units interspersed with regulatory elements). Viewed in this light, transgenes display all the hallmarks of selfish genetic elements and are thus recognized and targeted for silencing by host cells.

That we can take non-imprinted sequences (vector backbone + GFP + *C. elegans* genes) and derive an imprinted transgene (from an organism not known to imprint endogenous genes) was rather unexpected. Like ourselves, at least two other groups have fortuitously created imprinted transgenes from non-imprinted source sequences in different systems (KEARNS *et al.* 2000; MARTIN and MCGOWAN 1995; SWAIN *et al.* 1987). Experiments of this nature raise the intriguing question of whether imprinted transgenes can be methodically constructed from certain parameters. With our current understand of imprinting mechanisms, the answer at present is a "no", at least not entirely from rational design. Certain non-imprinted genes, such as the β-globin locus, can be made to express in a parent-of-origin manner by insertion of the DMR from the *H19* locus (TANIMOTO *et al.* 2005). But the chimera transgene does not completely recapitulate the endogenous *H19* imprint, as the imprint in the chimera appears to be established after fertilization, rather than during gametogenesis (TANIMOTO *et al.* 2005). In another experiment, a sequence from within the *H19* DMR acquired a silencing function in *Drosophila* (LYKO *et al.* 1997), although expression of the reporter gene was

not dependent upon parent-of-origin. Numerous other experiments, mainly in mice, have confirmed the partial modularity of DMRs. Thus, our own experiments, in conjunction with that of others described in this paragraph, seem to suggest that given some (as yet unknown) sequence and/or structural features, some imprinted effects can be created from non-imprinted sources, independent of genomic location (as the extra-chromosomal *ccEx3815* is also imprinted). This hypothesis would predict that *ccIn3861*, for which we have shown to be structurally similar to the imprinted *ccEx3815*, *ccIn3852*, and *ccIn3862* transgene arrays, to be also imprinted. Although we have not observed *ccIn3861* to have a parent-of-origin effect, it is possible that a very subtle imprinting effect existed in *ccIn3861* that is below the detection threshold of our assay, which has an estimated error of 15%. One possibility would be that certain loci might generate a parent-of-origin effect only when genetically "pushed" in a specific direction. Anecdotal evidence with *ccIn3861* suggests that this may be a productive research direction: an experiment analogous to that in Chapter 3 Figure 3.10 was performed in which *ccIn3861* was transmitted for 10 consecutive generations through either the sperm or oocyte lineage. Although there was no detectable change in *ccIn3861* expression between the first and the tenth generation sperm-derived array, there was a statistically significant decrease (30%) in expression levels between the first and tenth generation oocyte-transmitted array.

We note that in contrast to imprinting in mammals and especially the mealybugs, where the imprinted locus is either on or off, the imprinting effects we have observed in our set of transgenes are modest, but can be clearly discerned by assaying populations (rather than individuals). The biology of mealybugs dictates that imprinting in these

insects should be strictly qualitative. Mealybugs do not have a sex chromosome (BONGIORNI and PRANTERA 2003). Rather, sex determination is dictated by genome dosage: embryos in which the entire paternal genome is heterochromatinized and eliminated develop into males (BONGIORNI and PRANTERA 2003; BROWN and CHANDRA 1977; KHOSLA *et al*. 2006). In imprinting of individual genes, as is the case in mammals and plant, we believe that there is no reason why imprinting should only be qualitative in nature. Increasingly, as assays to detect gene expression have become more sensitive, more and more cases of quantitative differences (rather than simple "on-off" situations) of imprinted gene expression have been documented. Some imprinted genes previously thought to display qualitative differential expression have now been observed to have quantitative differential expression. In such cases, the "silenced" allele is expressed at a very low level or only in a subset of tissues (AINSCOUGH *et al*. 2000; CURCHOE *et al*. 2005; DELTOUR *et al*. 1995; HU *et al*. 1998; ISHIHARA *et al*. 2000; LATHAM 1995; LAU *et al*. 2004; LEE *et al*. 1997; LIU *et al*. 2005; LOVISETTI-SCAMIHORN *et al*. 1999; NISHIWAKI *et al*. 1997; SAKAMOTO *et al*. 2004; VERONA and BARTOLOMEI 2004; WILLIAMSON *et al*. 2004; WYLIE *et al*. 2003; YEVTODIYENKO *et al*. 2004).

Not only does *ccIn3862* express in a parent-of-origin manner, we have shown that the stability of its epigenetic state is a function of the number of consecutive generations it is kept in the male or female germline (Chapter 3, Figure 3.10C-D). That is, the stability of its imprint can be modulated by the length of time it is kept in a particular germline. This suggests that each generation is a limited window of opportunity in which the oocyte or sperm can lay down its gamete-specific marks.

An intriguing question is the extent to which endogenous genes might be imprinted, albeit very subtly. This can be investigated genetically by taking advantage of semi-dominant mutations in which heterozygotes show an intermediate phenotype between wildtype and homozygous mutants. That is, the severity of the phenotype increases in the order $\frac{+}{+} < \frac{m_{SD}}{+} < \frac{m_{SD}}{m_{SD}} \approx \frac{m_{SD}}{Df}$, where m_{SD} is a semi-dominant mutation and *Df* is a deficiency. Preliminary investigations of two semi-dominant loci, *unc-54*(*e1152*) and *dpy-25*(*e817sd*), suggest that this also may be a viable means to investigate endogenous imprinting. By transmitting the wildtype chromosome through either the male or female germline and assaying the resulting F1 animals, we should, in principle, be able to observe differences in phenotype (if it existed). In preliminary experiments with *unc-54*(*e1152*) and *dpy-25*(*e817sd*), we have not observed any differences with these loci. It is possible that differences existed, but were too subtle to be detected. Numerous other dominant and semi-dominant alleles are available in the *C. elegans* community that can be subjected to similar tests (PARK and HORVITZ 1986).

A second approach to investigating imprinting of endogenous genes is to measure the relative mRNA levels of a marker gene resulting from reciprocal crosses. A good candidate in which to investigate is *unc-54*, since the promoter of this gene was used to drive GFP expression in *ccEx3815* and its derivatives. Multiple single nucleotide polymorphisms (SNP) exist between the Bristol and CB4856 (Hawaiian) strain of *C. elegans* for the *unc-54* gene. The experiment would entail performing reciprocal crosses, reverse transcribing a polymorphic section of the *unc-54* message of F1 animals, and performing PCR to amplify the products. Since the two alleles differ by only a single

nucleotide, the kinetics of PCR amplification would be virtually identical for both alleles. Thus, the PCR needs to be quantitative on a relative (but not an absolute) level.

Needs for additional tools to study chromatin in *C. elegans*

Chromatin research conducted in other systems such as yeast, *Drosophila*, plants, and mammals has greatly expanded our understanding of the mechanisms by which chromatin modulates epigenetic activities. The mechanisms by which chromatin operate appear to be fundamental, and can be abstracted across widely differing taxa. Chromatin research in *C. elegans* has not matured to the level that it has in other model systems, however, primarily due to difficulties in isolating nuclei from a single tissue. Fortunately, many paradigms learned from other systems are applicable to *C. elegans*.

We initiated the development of a tool with the hope of allowing us to systematically study bulk chromatin in *C. elegans*. In particular, we engineered transgenic lines expressing the *E. coli* dam methyltransferase in muscle tissue, which methylates adenine in the target sequence GATC, in a single tissue type. Because *C. elegans* does not possess endogenous DNA methylation, the methyltransferase would not likely be subject to regulation in *C. elegans*. This is an advantage that allows us to ask which regions of the genome are susceptible to DNA methylation (and thus presumably less heterochromatic) and potentially to mark chromatin derived from a single cell type. Other groups have successfully employed similar strategies to probe chromatin structure in representative model organisms (BOIVIN and DURA 1998; BURYANOV and SHEVCHUK 2005; VAN STEENSEL *et al*. 2001; WINES *et al*. 1996).

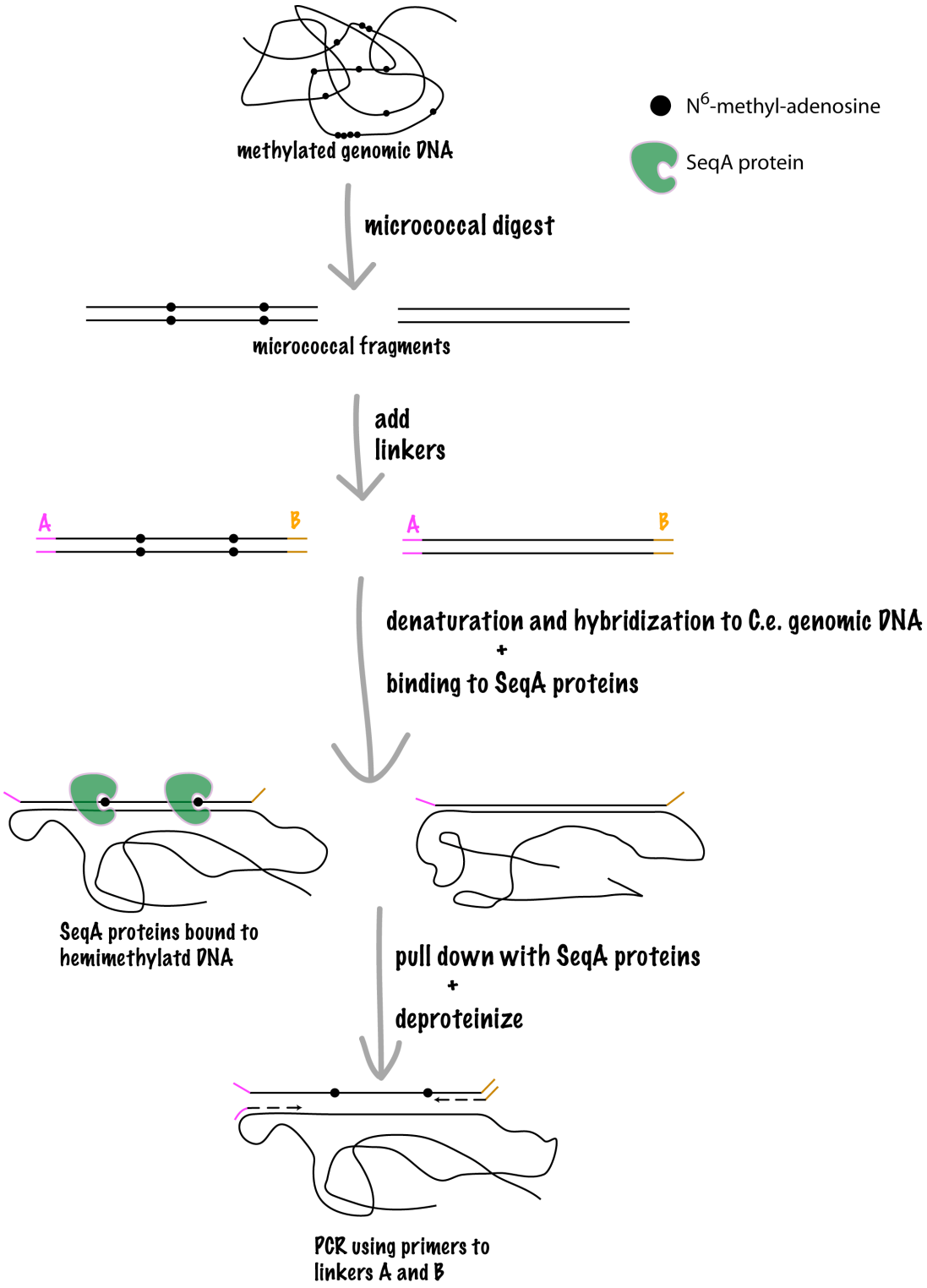
Preliminary data indicate that the targets of dam methyltransferase appear to be global. Additionally, from within muscle cells, dam methyltransferase appears to have

access to all categories of genes. With only 0.11% of the total potential target sites examined, we cannot make a general statement as to which regions are accessible (and thus presumably euchromatic) and which regions are inaccessible to dam methyltransferase. A large cloning effort examining about 100,00 target sites (≈ 30% of potential target sites in the genome) would allow higher resolution mapping of dam methyltransferase target sites. Alternatively, there are *E. coli* regulatory proteins that recognize N^6-methyl-adenine. One such protein is SeqA, which binds hemi-methylated GATC (HAN *et al*. 2004; LOBNER-OLESEN *et al*. 2005). Since muscle cells in *C. elegans* are post-mitotic, visualization of SeqA::GFP foci *in vivo* might be very rare. Instead, SeqA can be used to specifically isolate methylated fragments (after micrococcal nuclease digestion) for sequencing. The methodology is shown in Figure 5.1. Genomic (methylated) DNA from DAM-expressing animals are digested with micrococcal nuclease, yielding nucleosomal fragments of ≈146bp in length. A fraction of the total micrococcal-derived fragments will be methylated. Unique linkers A and B are ligated to both ends of the micrococcal fragments. The fragments are denatured and hybridized to wildtype (non-methylated) *C. elegans* genomic DNA. Micrococcal fragments that are methylated will form hemi-methylated hybrids with the wildtype DNA, which can be isolated by binding to SeqA and separated from non-methylated hybrids. Isolated hemi-methylated hybrids are subjected to PCR using primers to linkers A and B, followed by sequencing. We believe this procedure will allow us to greatly increase the resolution of DAM target sites in the genome and allow us to determine regions of DAM exclusion (i.e. heterochromatin) in the *C. elegans* genome.

FIGURES

Figure 5.1. Tissue-specific analysis of chromatin accessibility using dam methyltransferase and SeqA. The procedure employs the *E. coli* SeqA protein to isolate methylated targets of dam methyltransferase followed by sequencing of the isolated fragment. Filled circles represent N^6-methyl-adenosine. A (purple) and B (orange) are unique linkers ligated to micrococcal fragments. SeqA proteins are represented by green pac-men.

Figure 5.1


methylated genomic DNA
N6-methyl-adenosine
SeqA protein
micrococcal digest
micrococcal fragments
add
linkers
A
B
A
B
denaturation and hybridization to C.e. genomic DNA
+
binding to SeqA proteins
SeqA proteins bound to
hemimethylatd DNA
pull down with SeqA proteins
+
deproteinize
PCR using primers to
linkers A and B

REFERENCES

AINSCOUGH, J. F., R. M. JOHN, S. C. BARTON and M. A. SURANI, 2000 A skeletal muscle-specific mouse Igf2 repressor lies 40 kb downstream of the gene. Development **127:** 3923-3930.

ARAVIN, A. A., N. M. NAUMOVA, A. V. TULIN, V. V. VAGIN, Y. M. ROZOVSKY *et al.*, 2001 Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. Curr Biol **11:** 1017-1027.

ASSAAD, F. F., K. L. TUCKER and E. R. SIGNER, 1993 Epigenetic repeat-induced gene silencing (RIGS) in Arabidopsis. Plant Mol Biol **22:** 1067-1085.

BOIVIN, A., and J. M. DURA, 1998 In vivo chromatin accessibility correlates with gene silencing in Drosophila. Genetics **150:** 1539-1549.

BONGIORNI, S., and G. PRANTERA, 2003 Imprinted facultative heterochromatization in mealybugs. Genetica **117:** 271-279.

BROWN, S. W., and H. S. CHANDRA, 1977 *Chromosome imprinting and the differential regulation of homologous chromosomes*. Academic Press, New York.

BURYANOV, Y., and T. SHEVCHUK, 2005 The use of prokaryotic DNA methyltransferases as experimental and analytical tools in modern biology. Anal Biochem **338:** 1-11.

CHUNG, J. H., M. WHITELEY and G. FELSENFELD, 1993 A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. Cell **74:** 505-514.

CHUNG, J. H., A. C. BELL and G. FELSENFELD, 1997 Characterization of the chicken beta-globin insulator. Proc Natl Acad Sci U S A **94:** 575-580.

CURCHOE, C., S. ZHANG, Y. BIN, X. ZHANG, L. YANG *et al.*, 2005 Promoter-specific expression of the imprinted IGF2 gene in cattle (Bos taurus). Biol Reprod **73:** 1275-1281.

DAY, C. D., E. LEE, J. KOBAYASHI, L. D. HOLAPPA, H. ALBERT *et al.*, 2000 Transgene integration into the same chromosome location can produce alleles that express at a predictable level, or alleles that are differentially silenced. Genes Dev **14:** 2869-2880.

DELTOUR, L., X. MONTAGUTELLI, J. L. GUENET, J. JAMI and A. PALDI, 1995 Tissue- and developmental stage-specific imprinting of the mouse proinsulin gene, Ins2. Dev Biol **168:** 686-688.

DORER, D. R., and S. HENIKOFF, 1994 Expansions of transgene repeats cause heterochromatin formation and gene silencing in Drosophila. Cell **77:** 993-1002.

FIRE, A., D. ALBERTSON, S. W. HARRISON and D. G. MOERMAN, 1991 Production of antisense RNA leads to effective and specific inhibition of gene expression in *C. elegans* muscle. Development **113:** 503-514.

GARRICK, D., S. FIERING, D. I. MARTIN and E. WHITELAW, 1998 Repeat-induced gene silencing in mammals. Nat Genet **18:** 56-59.

GDULA, D. A., T. I. GERASIMOVA and V. G. CORCES, 1996 Genetic and molecular analysis of the gypsy chromatin insulator of Drosophila. Proc Natl Acad Sci U S A **93:** 9378-9383.

GRISHOK, A., H. TABARA and C. C. MELLO, 2000 Genetic requirements for inheritance of RNAi in C. elegans. Science **287:** 2494-2497.

GRISHOK, A., J. L. SINSKEY and P. A. SHARP, 2005 Transcriptional silencing of a transgene by RNAi in the soma of C. elegans. Genes Dev **19:** 683-696.

GRUMMT, I., and C. S. PIKAARD, 2003 Epigenetic silencing of RNA polymerase I transcription. Nat Rev Mol Cell Biol **4:** 641-649.

HAN, J. S., S. KANG, S. H. KIM, M. J. KO and D. S. HWANG, 2004 Binding of SeqA protein to hemi-methylated GATC sequences enhances their interaction and aggregation properties. J Biol Chem **279:** 30236-30243.

HIROCHIKA, H., H. OKAMOTO and T. KAKUTANI, 2000 Silencing of retrotransposons in arabidopsis and reactivation by the ddm1 mutation. Plant Cell **12:** 357-369.

HSIEH, J., and A. FIRE, 2000 Recognition and silencing of repeated DNA. Annu Rev Genet **34:** 187-204.

HU, J. F., H. ORUGANTI, T. H. VU and A. R. HOFFMAN, 1998 Tissue-specific imprinting of the mouse insulin-like growth factor II receptor gene correlates with differential allele-specific DNA methylation. Mol Endocrinol **12:** 220-232.

ISHIHARA, K., N. HATANO, H. FURUUMI, R. KATO, T. IWAKI *et al*., 2000 Comparative genomic sequencing identifies novel tissue-specific enhancers and sequence elements for methylation-sensitive factors implicated in Igf2/H19 imprinting. Genome Res **10:** 664-671.

JENSEN, S., M. P. GASSAMA and T. HEIDMANN, 1999 Taming of transposable elements by homology-dependent gene silencing. Nat Genet **21:** 209-212.

JEONG BR, B. R., D. WU-SCHARF, C. ZHANG and H. CERUTTI, 2002 Suppressors of transcriptional transgenic silencing in Chlamydomonas are sensitive to DNA-

damaging agents and reactivate transposable elements. Proc Natl Acad Sci U S A **99:** 1076-1081.

KEARNS, M., J. PREIS, M. MCDONALD, C. MORRIS and E. WHITELAW, 2000 Complex patterns of inheritance of an imprinted murine transgene suggest incomplete germline erasure. Nucleic Acids Res **28:** 3301-3309.

KETTING, R. F., T. H. HAVERKAMP, H. G. VAN LUENEN and R. H. PLASTERK, 1999 Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell **99:** 133-141.

KHOSLA, S., G. MENDIRATTA and V. BRAHMACHARI, 2006 Genomic imprinting in the mealybugs. Cytogenet Genome Res **113:** 41-52.

KUNERT, N., J. MARHOLD, J. STANKE, D. STACH and F. LYKO, 2003 A Dnmt2-like protein mediates DNA methylation in Drosophila. Development **130:** 5083-5090.

LABRADOR, M., and V. G. CORCES, 2002 Setting the boundaries of chromatin domains and nuclear organization. Cell **111:** 151-154.

LATHAM, K. E., 1995 Stage-specific and cell type-specific aspects of genomic imprinting effects in mammals. Differentiation **59:** 269-282.

LAU, J. C., M. L. HANEL and R. WEVRICK, 2004 Tissue-specific and imprinted epigenetic modifications of the human NDN gene. Nucleic Acids Res **32:** 3376-3382.

LEE, M. P., R. J. HU, L. A. JOHNSON and A. P. FEINBERG, 1997 Human KVLQT1 gene shows tissue-specific imprinting and encompasses Beckwith-Wiedemann syndrome chromosomal rearrangements. Nat Genet **15:** 181-185.

LI, Q., M. ZHANG, H. HAN, A. ROHDE and G. STAMATOYANNOPOULOS, 2002 Evidence that DNase I hypersensitive site 5 of the human beta-globin locus control region

functions as a chromosomal insulator in transgenic mice. Nucleic Acids Res **30:** 2484-2491.

LIPPMAN, Z., A. V. GENDREL, M. BLACK, M. W. VAUGHN, N. DEDHIA *et al.*, 2004 Role of transposable elements in heterochromatin and epigenetic control. Nature **430:** 471-476.

LIU, J., M. CHEN, C. DENG, D. BOURC'HIS, J. G. NEALON *et al.*, 2005 Identification of the control region for tissue-specific imprinting of the stimulatory G protein alpha-subunit. Proc Natl Acad Sci U S A **102:** 5513-5518.

LOBNER-OLESEN, A., O. SKOVGAARD and M. G. MARINUS, 2005 Dam methylation: coordinating cellular processes. Curr Opin Microbiol **8:** 154-160.

LOVISETTI-SCAMIHORN, P., R. FISCHER-COLBRIE, B. LEITNER, G. SCHERZER and H. WINKLER, 1999 Relative amounts and molecular forms of NESP55 in various bovine tissues. Brain Res **829:** 99-106.

LYKO, F., J. BRENTON, M. A. SURANI and R. PARO, 1997 An imprinting element from the mouse H19 locus functions as a silencer in Drosophila. Nat Genet **16:** 171-173.

LYON, M. F., 2003 The Lyon and the LINE hypothesis. Semin Cell Dev Biol **14:** 313-318.

MARTIN, C. C., and R. A. MCGOWAN, 1995 Parent-of-origin specific effects on the methylation of a transgene in zebrafish. Devel. Genet. **17:** 233-239.

MCBURNEY, M. W., T. MAI, X. YANG and K. JARDINE, 2002 Evidence for repeat-induced gene silencing in cultured Mammalian cells: inactivation of tandem repeats of transfected genes. Exp Cell Res **274:** 1-8.

MELLO, C. C., J. M. KRAMER, D. STINCHCOMB and V. AMBROS, 1991 Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences. Embo J **10:** 3959-3970.

NISHIWAKI, K., N. NIIKAWA and M. ISHIKAWA, 1997 Polymorphic and tissue-specific imprinting of the human Wilms tumor gene, WT1. Jpn J Hum Genet **42:** 205-211.

NOLAN, T., L. BRACCINI, G. AZZALIN, A. DE TONI, G. MACINO *et al.*, 2005 The post-transcriptional gene silencing machinery functions independently of DNA methylation to repress a LINE1-like retrotransposon in Neurospora crassa. Nucleic Acids Res **33:** 1564-1573.

PAL-BHADRA, M., B. A. LEIBOVITCH, S. G. GANDHI, M. RAO, U. BHADRA *et al.*, 2004 Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. Science **303:** 669-672.

PALMITER, R. D., and R. L. BRINSTER, 1986 Germ-line transformation of mice. Annu Rev Genet **20:** 465-499.

PARK, E. C., and H. R. HORVITZ, 1986 Mutations with dominant effects on the behavior and morphology of the nematode Caenorhabditis elegans. Genetics **113:** 821-852.

PELISSIER, T., and M. WASSENEGGER, 2000 A DNA target of 30 bp is sufficient for RNA-directed DNA methylation. Rna **6:** 55-65.

SAKAMOTO, A., J. LIU, A. GREENE, M. CHEN and L. S. WEINSTEIN, 2004 Tissue-specific imprinting of the G protein Gsalpha is associated with tissue-specific differences in histone methylation. Hum Mol Genet **13:** 819-828.

SANTORO, R., J. LI and I. GRUMMT, 2002 The nucleolar remodeling complex NoRC mediates heterochromatin formation and silencing of ribosomal gene transcription. Nat Genet **32:** 393-396.

SELKER, E. U., 1990 Premeiotic instability of repeated sequences in Neurospora crassa. Annu Rev Genet **24:** 579-613.

SELKER, E. U., 1997 Epigenetic phenomena in filamentous fungi: useful paradigms or repeat-induced confusion? Trends Genet **13:** 296-301.

SIJEN, T., and R. H. PLASTERK, 2003 Transposon silencing in the Caenorhabditis elegans germ line by natural RNAi. Nature **426:** 310-314.

STINCHCOMB, D. T., J. E. SHAW, S. H. CARR and D. HIRSH, 1985 Extrachromosomal DNA transformation of *Caenorhabditis elegans*. Mol Cell Biol **5:** 3484-3496.

SWAIN, J. L., T. A. STEWART and P. LEDER, 1987 Parental legacy determines methylation and expression of an autosomal transgene: a molecular mechanism for parental imprinting. Cell **50:** 719-727.

TABARA, H., M. SARKISSIAN, W. G. KELLY, J. FLEENOR, A. GRISHOK *et al.*, 1999 The rde-1 gene, RNA interference, and transposon silencing in C. elegans. Cell **99:** 123-132.

TAJIMA, S., K. SHINOHARA, M. FUKUMOTO, R. ZAITSU, J. MIYAGAWA *et al.*, 2006 Ars insulator identified in sea urchin possesses an activity to ensure the transgene expression in mouse cells. J Biochem (Tokyo) **139:** 705-714.

TANIMOTO, K., M. SHIMOTSUMA, H. MATSUZAKI, A. OMORI, J. BUNGERT *et al.*, 2005 Genomic imprinting recapitulated in the human beta-globin locus. Proc Natl Acad Sci U S A **102:** 10250-10255.

UDVARDY, A., E. MAINE and P. SCHEDL, 1985 The 87A7 chromomere. Identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains. J Mol Biol **185:** 341-358.

VAN STEENSEL, B., J. DELROW and S. HENIKOFF, 2001 Chromatin profiling using targeted DNA adenine methyltransferase. Nat Genet **27:** 304-308.

VASTENHOUW, N. L., S. E. FISCHER, V. J. ROBERT, K. L. THIJSSEN, A. G. FRASER *et al.*, 2003 A genome-wide screen identifies 27 genes involved in transposon silencing in C. elegans. Curr Biol **13:** 1311-1316.

VERONA, R. I., and M. S. BARTOLOMEI, 2004 Role of H19 3' sequences in controlling H19 and Igf2 imprinting and expression. Genomics **84:** 59-68.

WHITELAW, E., and D. I. MARTIN, 2001 Retrotransposons as epigenetic mediators of phenotypic variation in mammals. Nat Genet **27:** 361-365.

WILLIAMSON, C. M., S. T. BALL, W. T. NOTTINGHAM, J. A. SKINNER, A. PLAGGE *et al.*, 2004 A cis-acting control region is required exclusively for the tissue-specific imprinting of Gnas. Nat Genet **36:** 894-899.

WINES, D. R., P. B. TALBERT, D. V. CLARK and S. HENIKOFF, 1996 Introduction of a DNA methyltransferase into Drosophila to probe chromatin structure in vivo. Chromosoma **104:** 332-340.

WOODHOUSE, M. R., M. FREELING and D. LISCH, 2006 The mop1 (mediator of paramutation1) mutant progressively reactivates one of the two genes encoded by the MuDR transposon in maize. Genetics **172:** 579-592.

WU-SCHARF, D., B. JEONG, C. ZHANG and H. CERUTTI, 2000 Transgene and transposon silencing in Chlamydomonas reinhardtii by a DEAH-box RNA helicase. Science **290:** 1159-1162.

WYLIE, A. A., D. J. PULFORD, A. J. MCVIE-WYLIE, R. A. WATERLAND, H. K. EVANS *et al.*, 2003 Tissue-specific inactivation of murine M6P/IGF2R. Am J Pathol **162:** 321-328.

YE, F., and E. R. SIGNER, 1996 RIGS (repeat-induced gene silencing) in Arabidopsis is transcriptional and alters chromatin configuration. Proc Natl Acad Sci U S A **93:** 10881-10886.

YEVTODIYENKO, A., E. Y. STESHINA, S. C. FARNER, J. M. LEVORSE and J. V. SCHMIDT, 2004 A 178-kb BAC transgene imprints the mouse Gtl2 gene and localizes tissue-specific regulatory elements. Genomics **84:** 277-287.

CURRICULUM VITAE

Date of Birth

August 25, 1974

Education

1981-1982	Mountain View Elementary School, Olympia, WA
1982-1989	Repetto Elementary School, Monterey Park, CA
1989-1993	Mark Keppel High School, Alhambra, CA
1993-1998	B.A. – Biochemistry and Molecular Biology: University of California, Berkeley
1993-1998	B.S. – Plant Biology: University of California, Berkeley
2000-2006	Ph.D. – Biology: Johns Hopkins University

Publications

Sha K. and Fire A. (2005) Imprinting Capacity of Gamete Lineages in *Caenorhabditis elegans*. Genetics 170:1633-1652

Kim I., Hempel F.D., **Sha K**, Pfluger J., Zambryski P. (2002) Identification of a developmental transition in plasmodesmatal function during embryogenesis in *Arabidopsis thaliana*. Development 129:1261-1272

Academic Awards and Honors

- Platform and Poster Presentation (2005): 13th International C. elegans Meeting, University of California, Los Angeles
- Poster Presentation (2004): International Conference on Genomic Imprinting, Oxford University, England
- Dean's Teaching Fellowship (2003): Johns Hopkins University
- Dupont Teaching Award (2001): Johns Hopkins University